

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



WELTORGANISATION FÜR GEISTIGES EIGENTUM

Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



(51) Internationale Patentklassifikation 6 : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027		A2	(11) Internationale Veröffentlichungsnummer: WO 99/33998 (43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)
(21) Internationales Aktenzeichen: PCT/EP98/08216 (22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98) (30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE (71) Anmelder (für alle Bestimmungsstaaten ausser US): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE). (72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggen- dorfstrasse 59, D-51061 Köln (DE). (74) Gemeinsamer Vertreter: BAYER AKTIENGE- SELLSCHAFT; D-51368 Leverkusen (DE).		(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  Veröffentlicht Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.	
(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF (54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOM- ERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG			
<p>RT MOTIF 8 27-Motif 1-3 4 5 6 7 MOTIF 1 START A. FULL LENGTH DEL1 DEL2 DEL3 DEL1+2 INS1 INS2 INS3 INS4 A. FULL LENGTH</p>			
(57) Abstract <p>The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.</p>			

**(57) Zusammenfassung**

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

**LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen  
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-  
dung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten  
10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der  
30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt



werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

#### Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

30

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

#### 10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34<sup>+</sup>38<sup>+</sup>-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

5       Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomerasequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase  
10       enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomerasequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.  
15

Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweiligen Strahlen schädigen nicht nur die Tumorzellen,  
20       sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgewebe  
25       wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden  
30       dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

#### Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein“) spezifisch gebunden werden.

20

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

25

30

5  
10  
15  
20  
25  
30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon  
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere  
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).  
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ , TNF, TNF- $\alpha$ , TNF- $\beta$ , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls  
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 $\alpha$ , - $\beta$ ), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle  $\beta$ -Glukuronidase, pflanzliche  $\beta$ -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle  $\beta$ -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC),  $\beta$ -Galaktosidase ( $\beta$ -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH),  $\beta$ -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch  
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch  
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,  
15 zen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.  
20

Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.  
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-



turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF  $\beta$ , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

30

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen  
5 Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen,  
15 die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der  
20 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:  
30

- 5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reporter-  
gen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reporter-Genaktivität, um einen diagnostischen Wert zu erhalten;
- 10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reporter-Genkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten  $\lambda$ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene  $\lambda$ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des  
10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei  $\lambda$ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten  $\lambda$ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

15 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene  $\lambda$ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur  
20 14: nicht verdaut.

B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

30

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTTAAGTTGTAGCTTAACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

20

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymchnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

30

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

30

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

15 Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

20

25 Tab. 1: Exon Intron Übergänge des hTC-Gens

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

30

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in  
10 Intron 2.



Tab. 1

3' Acceptor Sequence			5' Donor Sequence					
Intron	Exon	Exon No.	bp	Exon	Intron	Intron on No.	bp	
5' flankierende Region								
cagggcgctcccccgag	GTTCAGGCAGCGTGCGT	1	281	CGCCCCCTCCTTCGCCAG	gtggcctccccgggtcg	1	104	
catgtccttctcgttttaag	GTGTCTGCCTGNAAGGAGC	2	1354	TGGCTGCGCAGGAGCCAG	gtgaggaggtggtggcgt	2	8616	
gaggggctctctatggag	GGGTGGCTGTGTTCCGGC	3	196	TGCAAAAGCATTGGAATCAG	gtactgtatccccacgcca	3	2089	
cccatgtgtcccccgag	ACAGCACATTGAAGAGGGTG	4	181	GTTCCGCAGAGAAAGAGG	gtgctgtgcttttggttta	4	687	
ctcgctccactcacaag	GCCGAGCGTCTCACCTCGA	5	180	TGAGCTGTACTTTTGTCAAG	gtgggtgcccggggaccccc	5	494	
ccctctcctctgcggag	GTGGATGTGACGGGGCGCGT	6	156	CAAGGCCCTTCAAGAGCCAC	gtaaggttcacgtgtgata	6	>4660	18
ctcccgctcgttttcgag	GTCTCTACCTTGACAGACC	7	96	TGCCGTGTCATCGAGCAG	gtctggggcaactgcccgtga	7	980	
ctgtgtcttccccgcccag	AGCTCTCCCTGNAATGAGG	8	86	CCGTGCGCATCAGGGGCAA	gtgagtcaggtggccaggt	8	2484	
gtattttcccttattctag	GTCTCTCCTCGGTTTGCTG	9	114	CGGGGATTTCGGCGGACGG	gtgagggcctcctctctcccc	9	1984	
catgtccccctctgccttag	GACCTGTCTCGGTTTGCTG	10	72	ACGCGAAACCTTCCTCAG	gtgagggcccgtgccgtgtg	10	1871	
attccccctgtgtcttag	GACCTGTCTCGGTTTGCTG	11	189	TGCAGAGCGCACTACTCCAG	gtgagcgcaactggccgga	11	3804	
tctttcttgccgactctag	CTATGCCCGGACCTCCATC	12	127	CCTGTTTCTGGATTGTCAG	gtgagcaggctgagtgtca	12	880	
ctgtccgccatcctctctag	GTGAACAGCCTCCAGACGG	13	62	TCCTGCTGCAGGCGTACAG	gtgagccgcccaccaaagggg	13	3187	
agcctctgtttccccag	GTTCACGCATGTGTCTG	14	125	CTGAAAGCCCAAGACGACG	gtatgtgcaggtgcccgtgc	14	781	
tctgatttttgcccccgag	GGATGTCTGTGGGGGCGAA	15	138	CTGGGCTCACTCAGACAG	gcaagtggtggtgagggcc	15	536	
	CCCAGACGCAGCTGAGTCG	16	664	TTTTTCAGTTTGGAAAAA	3' flankierende Region			

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

### Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

#### Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

## 5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca  $1,5 \times 10^6$  Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa  
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml  
15 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit  $1,5 \times 10^6$  cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte  
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*  
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen  
30 Experimenten jeweils 1 bis  $1,5 \times 10^6$  Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach  $\lambda$ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

### **Beispiel 3**

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde  $\lambda$ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte  $\lambda$ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

### **Beispiel 4**

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

30

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10 Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen

20

25

30

kloniert.

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten  
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-  
10 GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten  
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch  
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.



## Contig1:

	ACTTGAGCCC	AAGAGTTC	GGCTACGGTG	AGCCATGATT	GCAACACCAC	ACGCCAGCCT	TGGTGACAGA	70
	ATGAGACCCT	GTCTCAAAAA	AAAAA	AATTGAAATA	ATATAAGCA	TCTTCTCTGG	CCACAGTGG	140
5	ACAAAACCAG	AAATCAACAA	CAAGAGGAAT	TTTGA	ATACAAACAC	ATGAAAATTA	AACAATATAC	210
	TTCTGAATGA	CCAGTGAGTC	AATGAAGAAA	TTAAAAAGGA	AATTGAAAAA	TTTATTTAAG	CAATATGATA	280
	CGGAACATA	ACCTCTCAA	ACCCACGGTA	TACAGCAAAA	GCAGTGCTAA	GAAGGAAGTT	TATAGCTATA	350
	AGCAGCTACA	TCAAAAAAGT	AGAAAAGCCA	GGCCAGTGG	CTCATGCCTG	TAATCCCAGC	ACTTTGGGAG	420
	GCCAAAGGGG	GCAGATCGCC	TGAGGTGAGG	AGTTTCGAGAC	CAGCCTGACC	AACACAGAGA	AACCTTGTCTG	490
10	CTACTAAAA	TACAAAATTA	GCTGGGCATG	GTGGCACATG	CCTGTAATCC	CAGCTACTCG	GGAGGCTGAG	560
	GCAGGATAAC	CGCTTGAACC	CAGGAGGTGG	AGGTTGCGGT	GAGCCGGGAT	TGCGCCATTG	GACTCCAGCC	630
	TGGGTAAACAA	GAGTGAAACC	CTGTCTCAAG	AAAAA	AAGTAGAAAA	ACTTAAAAAT	ACAACTTAAT	700
	GATGCACCTT	AAAGAACTAG	AAAAGCAAGA	GCAAACTAAA	CCTAAAAATTG	GTAAGAGAAA	AGAAATAATA	770
	AAGATCAGAG	CAGAAATAAA	TGAAACTGAA	AGATAACAAT	ACAAAAGATC	AACAAAATTA	AAAGTTGGTT	840
15	TTTTGAAAG	ATAAACAAAA	TTGACAAACC	TTTGCCCA	CTAAGAAAA	AGGAAAGAAG	ACCTAAATAA	910
	ATAAAGTCAG	AGATGAAAA	AGAGACATTA	CAACTGATAC	CACAGAAAT	CAAAGGATCA	CTAGAGGCTA	980
	CTATGAGCAA	CTGTACACTA	ATAAATTGAA	AAACCTAGAA	AAAATAGATA	AATTCCTAGA	TGCATACAA	1050
	TACCAAGAT	TGAACCATGA	AGAAATCCAA	AGCCCAACAA	GACCAATAAC	AATAATGGGA	TTAAAGCCAT	1120
	AATAAAGAGT	CTCCTAGCAA	AGAGAAGCCC	AGGACCCAA	GGCTTCCCTG	CTGGATTTTA	CCAATCATTT	1190
20	AAAGAGAAAT	GAATTCCAAT	CCTACTCAA	CTATTCTGAA	AAATAGAGGA	AAGAATACTT	CCAAACTCAT	1260
	TCTACATGGC	CAGTATTACC	CTGATTCCAA	AACCAGACAA	AAACACATCA	AAAAACAA	AACAAAAA	1330
	CAGAAAGAAA	GAAACTACA	GGCCAATATC	CCTGATGAAT	ACTGATACAA	AAATCCTCAA	AAAAACACTA	1400
	GCAAAACAAA	TTAAACAACA	CCTTCGAAG	ATCATTCTATT	GTGATCAAGT	GGGATTTATT	CCAGGGATGG	1470
	AAGGATGGTT	CAACATATGC	AAATCAATCA	ATGTGATACA	TCATCCCAAC	AAAAATGAAGT	ACAAAACTA	1540
25	TATGATTATT	TCACTTTATG	CAGAAAAGC	ATTTGATAAA	ATTCTGCACC	CTTCATGATA	AAAAACCTCA	1610
	AAAAACAGG	TATACAGAA	ACATACAGCC	CAGGCACAGT	GGCTCACACC	TGCGATCCCA	GCACCTGGG	1680
	AGGCCAAGGT	GGGATGATTG	CTTGGGCCCA	GGAGTTTGAG	ACTAGCCTGG	GCAACAAAT	GAGACCTGGT	1750
	CTACAAAAA	CTTTTTTAA	AAATTAGCCA	GGCATGATGG	CATATGCCTG	TAGTCCCAGC	TAGTCTGGAG	1820
	GCTGAGGTGG	GAGAATCACT	TAAGCCTAGG	AGGTCGAGGC	TGCAGTGAGC	CATGAACATG	TCACCTGACT	1890
30	CCAGCCTAGA	CAACAGAA	AGACCCCACT	GAATAAGAG	AAGGAGAGG	AGAAGGGAGA	AGGGAGGGAG	1960
	AAGGGAGGAG	GAGGAGAGG	AGGAGGTGGA	GGAGAAGTGG	AAGGGGAAGG	GGAAGGGAAA	GAGGAAGAAG	2030
	AAGAAACATA	TTTCAACATA	ATAAAGCCC	TATATGACAG	ACCGAGGTAG	TATTATGAGG	AAAACTGAA	2100
	AGCCTTTCT	CTAAGATCTG	GAAATGACA	AGGGCCCACT	TTCAACACTG	TGATTCAACA	TAGTACTAGA	2170
35	AGTCTAGCT	AGAGCAATCA	GATAAGAGAA	AGAAATAAAA	GGCATCCAAA	CTGGAAAGGA	AGAAGTCAAA	2240
	TTATCCTGTT	TGCAGATGAT	ATGATCTTAT	ATCTGGAAAA	GACTTAAGAC	ACCACTAAAA	AACTATTAGA	2310
	GCTGAAATTT	GGTACAGCAG	GATACAAAT	CAATGTACAA	AAATCAGTAG	TATTCTTATA	TTCCAACAGC	2380
	AAACATCTG	AAAAAGAAAC	CAAAAAAGCA	GCTACAAATA	AAATTAACA	GCTAGGAATT	AACCAAGAAA	2450
	GTGAAAGATC	TCTACATGA	AAACTATAAA	ATGTTGATAA	AAGAAATGA	AGAGGGCACA	AAAAAGAAA	2520
40	AGATATTCCA	TGTTTCATAGA	TTGGAAGAA	AAATCTGTT	AAATGTCCA	TACTACCCCA	AGCAATTTAC	2590
	AAATTCATG	CAATCCCTAT	TAAATACTA	ATGACGTTCT	TCACAGAAAT	AGAAGAAACA	ATTCTAAGAT	2660
	TTGTACAGAA	CCACAAAGA	CCCAAGTAG	CCTGACCAAA	AAGAACAACA	CTGGAAGCAT	TTGGAAGCAT	2730
	CACATTACCT	GACTTCAAT	TATACTACAA	AGCTATAGTA	ACCCAACTA	CATGGTACTG	GCATAAAAAC	2800
	AGATGAGACA	TGGACCAGAG	GAACAGATA	GAGAATCCAG	AAACAAATCC	ATGCATCTAC	AGTGAACCTA	2870
45	TTTTTGACAA	AGGTGCCAAG	AACATACTTT	GGGAAAAAGA	TAATCTCTTC	AATAAATGGT	GCTGGAGGAA	2940
	CTGGATATCC	ATATGCAAAA	TAACAATACT	AGAACTCTGT	CTCTACCAT	ATACAAAAGC	AAATCAAAAT	3010
	GGATGAAAGG	CTTAAATCTA	AAACCTCAAA	CTTTGCAACT	ACTAAAAGAA	AACACCCGAG	AAACTCTCCA	3080
	GGACATTGGA	GTGGGCAAG	ACTTCTTGAG	TAATTCCTCTG	CAGGCACAGG	CAACCAAGC	AAAAACAGAC	3150
	AAATGGGATC	ATATCAAGTT	AAAAAGCTTC	TGCCAGCAAA	AGGAAACAAT	CAACAAAGAG	AAGAGACAA	3220
50	CCACAGAAATG	GGAGAAATATA	TTTGCAAACT	ATTCATCTAA	CAAGGAATTA	ATAACCAAGTA	TATATAAGGA	3290
	GCTCAAACTA	CTCTATAAGA	AAAAACCTA	ATAAGCTGAT	TTTCAAAAT	AAGCAAAAGA	TCTGGGTAGA	3360
	CATTTCCTCA	AATAAGTCAT	ACAAATGGCA	AACAGGCATC	TGAAAATGTG	CTCAACACCA	CTGATCATCA	3430
	GAGAAATGCA	AATCAAAACT	ACTATGAGAG	ATCATCTCAT	CCCAGTTAAA	ATGGCTTTTA	TTCAAAAGAC	3500
	AGGCAATAAC	AAATGCCAGT	GAGGATGTGG	ATAAAGGAA	ACCCTTGGAC	ACTGTTGGTG	GGAAATGGAAA	3570
55	TTGCTACCAC	TATGGAGAAC	AGTTTGAAG	TTCTCAAAA	AACTAAAAAT	AAAGCTACCA	TACAGCAATC	3640
	CCATTGCTAG	GTATATACTC	CAAAAAAGGG	AATCAGTGTA	TCAACAAGCT	ATCTCCACTC	CCACATTTAC	3710
	TGCAGCACTG	TTCATAGCAG	CCAAGTTTG	GAAGCAACCT	CAGTGTCAT	CAACAGACGA	ATGGAAAAAG	3780
	AAAAATGGT	GCACATACAC	AATGGAGTAC	TACGCAGCCA	TAAAAAGAA	TGAGATCCTG	TCAGTTGCAA	3850
	CAGCATGGGG	GGCACTGGTC	AGTATGTTAA	GTGAAATAAG	CCAGGCACAG	AAAGACAAAC	TTTTCATGTT	3920
60	CTCCCTTACT	TGTGGGAGCA	AAAAATTA	CAATTGACAT	AGAAATAGAG	GAGAATGGTG	GTCTAGAGG	3990
	GGTGGGGGAC	AGGGTGACTA	GAGTCAACAA	TAATTTATTG	TATGTTTTAA	AATAACTAAA	AGAGTATAAT	4060
	TGGGTTGTTT	GTAACACAAA	GAAAGGATAA	ATGCTTGAAG	GTGACAGATA	CCCCATTAC	CCTGATGTGA	4130
	TTATTACACA	TTGTATGCCT	GTATCAAAAT	ATCTCATGTA	TGCTATAGAT	ATAAACCCCTA	CTATATTAAA	4200
	AATTAATTT	TTAATGGCCA	GGCAGGTTGG	CTCATGTCCG	TAATCCCAGC	ACTTTGGGAG	GGCAGGCGG	4270
65	GTGGATCACC	TGAGGTGAGG	AGTTTGAAG	CAGTCTGGCC	ACCATGATGA	AACCTGTCT	CTACTAAAGA	4340
	TACAAAAAT	AGCCAGGCGT	GGTGGCAGAT	ACCTGTAGTC	CCAACACTC	AGGAGGCTGA	GACAGGAGAA	4410
	TTGCTTGAAC	CTGGGAGGCG	GAGGTTGAG	TGAGCCGAGA	TCATGCCACT	GCACTGCAGC	CTGGGTGACA	4480
	GAGCAAGACT	CCATCTCAAA	ACAAAAACA	AAAAAGAA	ATTAATTTG	TAATTTTAT	GTACCGTATA	4550
	AATATATACT	CTACTATATT	AGAAGTTAAA	AATTAACA	ATTATAAAG	GTAATTAACC	ACTTAATCTA	4620
70	AAATAAGAAC	AATGTATGTG	GGGTTTCTAG	CTTCTGAAGA	AGTAAAGATT	ATGGCCACGA	TGGCAGAAAT	4690
	GTGAGGAGGG	AACAGTGAA	GTTACTGTG	TTAGACGCTC	ATACTCTCTG	TAAGTGACTT	AATTTTAAAC	4760
	AAAGACAGGC	TGGGAGAAGT	TAAAGAGGCA	TTCTATAAGC	CCTAAACAA	CTGCTAATAA	TGGTGAAAGG	4830
	TAATCTCTAT	TAATTACCAA	TAATTACAGA	TATCTCTAAA	ATCGAGCTGC	AGAATTGGCA	CGTCTGATCA	4900
	CACCGTCTCT	TCATTACAGG	TGCTTTTTTT	CTTGTGTGCT	TGGAGATTTT	CGATTGTGTG	TTCTGTGTTG	4970
75	GTTAAACTTA	ATCTGTATGA	ATCCTGAAAC	GAAAAATGTT	GGTGATTTC	TCCAGAGAA	TTAGAGTACC	5040
	TGGCAGGAAG	CAGGTGGCTC	TGTGGACCTG	AGCCACTTCA	ATCTTCAAGG	GTCTCTGGCC	AAGACCCAGG	5110

	TGCAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCGATGAG	AAGCCTGCCT	5180
	CGTTGGTGAG	CAGCGCATGA	AGTGCCCTTA	TTTACGCTTT	GCAAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
	AGGCGGCCAG	CGGGAATGCA	AGGAGTCAGA	AGCCTCCTGC	TCAAACCCAG	GCCAGCAGCT	ATGGGCGCCA	5320
5	CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTCA	AGGCACCTCG	AAGTATGGCT	TAAATCTTTT	TTTACCTGTA	5390
	AGCAGTGACC	AAGGTGTATT	CTGAGGGAAG	CTTGAGTTAG	GTGCTTCTT	TAAACAGAA	AGTCATGGAA	5460
	GCACCTTCT	CAAGGGAAAA	CCAGACGCC	GCTCTGCGGT	CATTTACCTC	TTTCTCTCT	CCCTCTCTTG	5530
	CCCTCGCGGT	TTCTGATCGG	GACAGAGTGA	CCCCCGTGG	GCTTCTCCGA	GCCCGTGCTG	AGGACCTCT	5600
	TGCAAGGGC	TCCACAGACC	CCCGCCTGG	AGAGAGGAGT	CTGAGCCTGG	CTTAATAACA	AACCTGGGATG	5670
10	TGGCTGGGGG	CGGACAGCGA	CGCGGGGATT	CAAGACTTA	ATTCATGAG	TAAATTCAC	CTTCCACAT	5740
	CCGAATGGAT	TTGGATTTTA	TCTTAATATT	TTCTTAAATT	TCATCAAATA	ACATTCAGGA	CTGCAGAAAT	5810
	CCAAAGGGCT	AAAACAGGAA	CTGAGCTATG	TTTGCCAAAG	TCCAAGGACT	TAATAACCAT	GTTCAGAGGG	5880
	ATTTTTTCGCC	CTAAGTACTT	TTTATTGGTT	TTTATAAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGAGG	5950
	AGAGGCGCTG	GCGGCAGGGC	TATGAGCACG	GCAGGGCCAC	CGGGGAGAGA	GTCCCGCGCC	TGGGAGGCTG	6020
15	ACAGCAGGAC	CACTGACCGT	CCTCCCTGGG	AGCTGCCACA	TTGGGCAACG	CGAAGGGCGG	CACGCTGGCT	6090
	GTGACTCAGG	ACCCCATACC	GGCTTCTCTG	GCCACCCAC	ACTAACCCAG	GAAGTCACGG	AGCTCTGAAC	6160
	CGGTGGAAAC	GAACATGACC	CTTGCTGGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	GTTCAGAGGG	6230
	GGAAATGGCC	ATGTAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCCCTC	CATCATTATT	CATCTTCACC	6300
	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAACTCAG	TACAACACC	6370
20	ACTCTTTTAC	TAGGCCACCA	GAGCACGGSC	CACACCCCTG	ATATATTAAG	AGTCCAGGAG	AGATGAGGCT	6440
	GCTTTTCAGCC	ACCAGGCTGG	GGTGACAACA	CGCGCTGAAC	AGTCTGTTC	TCTAGACTAG	TAGACCTGG	6510
	CAGGCACCTC	CCGAGATTCT	AGGGCCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCTT	CCGCTCCAG	GCCTCAGCTT	CTCCAGCAGC	6650
	TTCTTAAACC	CTGGGTGGGC	CGTGTTCAG	CGCTACTGTC	TCACCTGTCC	CACCTGTGCT	TGCTCTCAGG	6720
25	ACGTAGCTCG	CACGGTTCCT	CCTCACATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	CGGTGGAAGG	6790
	GAGGAGATTG	TGCGCTCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGCCCTCC	GATGCAGGTT	6860
	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCGCTGCTC	CTGTCTCTG	CGGGGGCTG	6930
	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGGTG	TCTCTGCCCG	CATGGGTCTC	7000
30	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCGAG	GGCGCTCTTG	GGAAATGCAA	CATTGGGTG	7070
	TGAAAGTAGG	AGTGCTGTCT	CTCACCTAGG	TTCCAGGGCA	CAGGCTGGG	GATGGAGCCC	CGCCAGGGA	7140
	CCCGCCCTTC	TCTGCCACG	ACTTCTCTG	CCCTCTCCCT	CTGGAACACA	GAGTGGCAGT	TTCACAAGC	7210
	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	CCTGGGAATT	7280
	CACGTGACTA	CGCACATCAT	TACACACTCT	CCGTCCACGA	CCGACCCCGC	CTGTTTTATT	TAAATAGCTA	7350
35	CAAAAGCAGG	AAATCCCTCG	TAAATGTCTC	TTTAAACAAAC	TGGTTAAACA	AACGGGTCCA	TCCGCACGGT	7420
	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	GAATTACGGT	7490
	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAG	AAAGAAATTC	ACCCCATGGC	7560
	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGGGCA	GCTGGGGGCT	ACTGCACGCA	CCTTTTACTA	7630
	AAGCCAGTTT	CCTGGTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGGATGGG	7700
40	GGAAACCCGA	GGCTGTGCCA	TCTTTGCCAT	GCCCGAGTGT	CCTGGGCAGG	ATAATGCTCT	AGAGATGCC	7770
	AGTCTCTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCGCG	GCCCCAGGGC	CTTTGCAGGT	GTGATCTCCG	7840
	TGAGGACCTT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCGAAAG	GTAATCCAGG	GGTCTCGGGA	7910
	AGAGGCGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCAGTCTGA	GGCTGAAAG	7980
	GGAGGGAGGG	CCTCGAGCCC	AGGCCTGCAA	CGCCTCCAG	AAGCTGGAAG	AAGCGGGGAA	GGGACCTCTC	8050
45	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	TCCGGCTCTC	8120
	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTCAGAG	GCAACAGGAA	8190
	ACCCATGCAC	TGTGAATCTA	GGATTATTTT	AAAACAAAGG	TTTACAGAAA	CATCCAAAGG	CAGGGCTGAA	8260
	GTGCCCTCCG	GCAAGGGCAG	GGCAGGCAGC	AGTGATTATA	TTTAGCTATT	TTATTTTATT	TACTTACTTT	8330
	CTGAGACAGA	GTTATGCTCT	TGTTGCCACG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	TCTCAACCTC	8400
50	CGTCTCCTGG	TTTCAAGCAA	TTCTCTGCTC	TCAGCCTCCC	AAGTAGCTGG	GATTTTCAGG	GTGCACCACC	8470
	ACACCCGGCT	AATTTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	TCTCAAAATC	8540
	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CATGCACTC	8610
	GGCCTATTTA	ACCATTTTAA	AATCTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGTTC	8680
	AATTTCCCTT	TTACTCAGGA	GTTACCTCTC	TTTGATATTT	TCTGTAATTC	TTCTGAGACT	GGGATACAC	8750
	CGTCTCTTGA	CATATTCACA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACTGCAG	GGGCAGCTGG	8820
55	GAGGCTGCAG	GCCTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	ATCAGGGGCG	8890
	AAGTGTGGAC	ACTGTCTCGA	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	AAGTCCATCC	8960
	CTCCTACTCT	ACTGGGATTG	AGCCCTTCTC	CTATCCCCC	CCAGGGGACG	AGGAGTTTCT	CTCACTCCTG	9030
	TGGAGGAAGG	AATGATACTT	TGTTATTTTT	CACCTGCTGG	ACTGAATCCA	CTGTTTCATT	TGTTGGTTTG	9100
	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTGTCTCA	GGCTGGAGGG	AGTGCAATGG	CGCGATCTTG	9170
60	GCTTTACTGA	GCCTCTCGCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	GCTGGGATTA	9240
	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGATTTTT	TAGTAGAGAC	GSGGGTGGGT	GGGGTTTACC	9310
	ATGTTGGCCA	GGCTGGTCTC	GAACTTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCTCT	AAAGTGCTGG	9380
	GATTACAGGT	GTGAGCCACC	ATGCCAGGCT	CAGAATTTAC	TCTGTTTGA	AACATCTGGG	TCTGAGGTAG	9450
	GAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAGCCAA	TGATAGAAAT	TTTTTATTGT	TGTTAGAACA	9520
65	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	TGCACCCATA	9590
	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	CCATGCACAT	9660
	GGTGTTAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCTT	TCTTCTCCCT	CTTTTAAAT	TGTGTTTTCT	9730
	ATGTTGGCTT	CTCTGCAGAG	AACCAGTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	TTTCCAAACC	9800
	GGCCCTTTGC	CCTAGTGCCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	ATCACTAAGG	9870
70	GGATTTCCTAG	AAGAGCGACC	TGTAATCCTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	GAGCGTGACA	9940
	GGCCAGGGAG	GGTGGGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAGTC	AATTTCTCTC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAGGTTACAT	TAAAGGTTGC	GTGTTTAGC	ATTTCAAGT	TTGCCGACCT	CAGCTACAGC	10080
	ATCCCTGCAA	GGCTCCTGGG	GACCCAGAAG	TTTCTCGCCC	CCTTAGATGC	AACTTGAGC	AACCCGGAGT	10150
	CTGGATTCTC	GGGAAGTCTC	CAGCTGTCTC	GGGTTGTGTC	CGGGGCCCA	GGTCTGGAGG	GGACCACTGG	10220
75	CGTGTGGGCT	TCTACTGCTG	GGCTGGAAAG	CGGGCTCTCT	AGCTCTGCAG	TCCGAGGCTT	GGAGCCAGGT	10290
	GGCTGGACCC	CGAGGCTGCC	CTCCACCTTG	TGCGGGCGGG	ATGTGACCAG	ATGTTGGGCT	CATCTGCCAG	10360
	ACAGAGTGCC	GGGGCCACGG	GTCAAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGGTGCAGC	GCCAGCAGGA	10430
	GGCCTGGGCT	CCATTTCCCA	CCCTTCTCTG	ACGGGACCGC	CCCGGTGGGT	GATTAACAGA	TTTGGGGTGG	10500

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCGCTG	TGTC AAGGAG	10570
	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACTCCGCGA	GGTCCCGCGT	GCCCGTCCAG	GGAGCAATGC	10640
	GTCTCTGGGT	TCGTCCCGAG	CCGCGTCTAC	GGCGCTCCGT	CCTCCCTTC	ACGTCCGGCA	TTCTGGTGC	10710
5	CCGGAGCCCG	ACGCCCCCGG	TCCGACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	GCGGCCAAG	10780
	GGTCGCCGCA	CGCACCTGT	CCCAGGGCCT	CCACATCATG	GCCCCCTCC	CGGGTTACCC	CACAGCCTAG	10850
	GCCGATTTCGA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCT	GGGAGCGCGA	GCGGCGCGCG	10920
	GGCGGGGAG	CGCGGGCCAG	ACCCCGGGGT	CCGCGCGGAG	CAGCTGCGCT	GTCGGGGCCA	GGCGGGGCTC	10990
	CCAGTGGATT	CGCGGGCACA	GACGCCAGG	ACCGCGCTCC	CCACGTGGCG	GAGGGACTGG	GGACCCGGCG	11060
10	ACCCGTCTCTG	CCCTTCACCC	TTCCAGCTCC	GCCTCTCTCC	CGCGGACCCC	GCCCCGTCCC	GACCCCTCCC	11130
	GGGTCCCCGG	CCCAGCCCCC	TCCGGGCCCT	CCCAGCCCC	CCCTTCCTT	TCCGCGGCCC	CGCCCTCTCC	11200
	TCCGCGCGCG	AGTTTCAGGC	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	GCGATGCGCG	GCGCTCCCCG	CTGCCGAGCC	GTGCGCTCCC	TGCTGCGCAG	CCACTACCGC	GAGGTGCTGC	11340
	CGCTGGCCAC	GTTCTGTCCG	CGCCTGGGGC	CCCAGGGCTG	GCGGTCTGGT	CAGCGCGGGG	ACCCGGCGGC	11410
15	TTTCGCGCGG	CTGCTGGCCC	AGTGCCTGGT	GTGCGTGCCC	TGGGACGCAC	GGCCGCCCCC	CGCCGCCCCC	11480
	TCCTTCCGCC	AGGTGGGCT	CCCCGGGGCT	GGCGTCCGGC	TGGGGTTGAG	GGCGGGCGGG	GGGAACCAAG	11550
	GACATGCGGA	GAGCAGCGCA	GGCGACTCAG	GGCGCTTCCC	CCGCAGGTGT	CCTGCTGAA	GAGGCTGGTG	11620
	GGCCGAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGCGAAGA	ACGTGCTGCG	CTTCGGGTTT	GCGCTGCTGG	11690
	ACGGGGCCCG	CGGGGGCCCC	CCCAGAGCCT	TCACCACCAAG	CGTGGCGCAG	TACCTGCCCA	GAGCGGTGAC	11760
20	CGACGCAGTG	CGGGGAGCGG	GGGCGTGGGG	GCTGCTGCTG	CGCGCGGTGG	GCGACGACGT	GCTGTTTCA	11830
	CTGCTGGCAC	GCTGCGCGCT	CTTTGTGCTG	GTGGCTCCCA	GCTGCGCTTA	CCAGGTGTGC	GGGCGCGCGC	11900
	TGTACCACTG	CGGCGCTGCC	ACTCAGGCCC	GGCCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGCGTCTGGG	11970
	ATGCGAAACGG	GCTTGAACCC	ATAGCGTCAG	GGAGGCCGGG	GTCCCCCTGG	GCCTGCCAGC	CCCGGGTGGC	12040
	AGGAGGCGCG	GGGCGAGTGC	CAGCCGAAGT	TGCCCGTTGC	CCAAGAGGCC	CAGGCGTGGC	GCTGCCCTGT	12110
25	AGCCGGAGCG	GACGCCCGTT	GGGCGAGGGT	CCTGGGCCCC	CCCGGGCAGG	ACCGGTGGAC	GAGGTGACCG	12180
	TGCTTTCTGT	GTGGTGTAC	CTGCCAGACC	CGCCGAAGAA	GCCACCTCTT	TGGAGGGTGC	GCTCTTGGC	12250
	ACGCGCCACT	CCACCCATC	CGTGGGCGCG	CAGCACCACG	CAGGCCCCCC	ATCCACATCG	CGGCCACCAC	12320
	GTCCCTGGGA	CACGCCCTGT	CCCCCGGTGT	ACGCCGAGAC	CAAGCACTTC	CTCTACTCCT	CAGGCGACAA	12390
	GGAGGAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCAGGCTGA	CTGGCGCTCG	GAGGCTCGTG	12460
30	GAGACCATCT	TTCTGGTTTC	CAGGCCCTGG	ATGCCAGGGA	CTCCCCGAG	GTGCCCCCGC	CTGCCCCAGC	12530
	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGCTGCTTGG	GAACCAACGC	CAGTGCCCTT	CGGGGGTGGT	12600
	CCTCAGACG	CAGTCCCGCG	TGCGAGCTGC	GGTCACCCCA	GCAGCCGGTG	TCTGTGCCCG	GGAGAAGCCC	12670
	CAGGGCTCTG	TGGCGGCCCG	CGAGGAGGAG	GACACAGACC	CCCGTCCGCT	GSTGCAGCTG	CTCCGCCAGC	12740
	ACAGCAGCCC	CTGGCAGGTG	TACGGCTTCG	TGCGGGCGCTG	CCTGCGCCCG	CTGGTGCCCC	CAGGCTCTGT	12810
35	GGGCTCCAGG	CACAAAGAAC	GCCGCTTCTT	CAGGAACACC	AAGAAAGTTA	TCTCCCTGGG	GAGGATGCC	12880
	AAGCTCTCGC	TGCAGGAGCT	AGCGTGGAA	GGGACTGCGC	TGGGCTGCGC	TTGGCTGCGC	AGGAGCCAGG	12950
	GTGAGGAGGT	GGTGGCGCTG	GAGGGCCAG	GCCCCAGAGC	TGAATGCAGT	AGGGGCTCAG	AAAGGGGGCG	13020
	AGCGAGAGCC	CTGGTCTCTC	TGTCTCATC	CTCACGTGGG	CACACGTGGC	TTTTCGCTCA	GGACCTCGAG	13090
	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGCATAA	ACTTACGAGG	TTACCTTCA	13160
40	CGTTTTGATG	GACACCGCGT	TTCCAGGCGC	CGAGGCCAGA	GCAGTGAACA	GAGGAGGCTG	GGCGGGCAG	13230
	TGGAGCCGGG	TTGCCGGCAA	TGGGAGAGAG	TGTTGGAAG	CACAGACGCT	CTGGCGAGGG	TGCTGCAAGG	13300
	TTACCTATAA	TCCTCTCTCG	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCC	CTCTCTCTGG	13370
	GGGTGGGAGG	TAAGGGTTTT	GCAGGTGCAC	GTGGTCAGCC	AATATGCAGG	TTTGTGTTTA	AGATTTAATT	13440
	GTGTGTTGAC	GGCCAGGTGC	GGTGGCTCAC	GCCGTAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TCACCTGAGG	TCAGGAGTTT	GAGACAGGCC	TGACCAACAT	GGTGAAACCC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGGTGG	TGTGTGCCGT	TAATCCCAGC	TACTTGGGAG	GCTGAGGCGA	GAGAATCACT	13650
	TGAACCCAGG	AGCGCGGAGC	TGCAGTGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	CGACAAGAGT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGT	CGTTGATTGT	GCCAGGACAG	GSTAGAGGGA	GAGGTGATAG	13790
	ACTGTTCTCC	AGCACAGATC	CTGGTCCCAT	CTTTAGGTAT	GAAGAGGGCC	ACATGGGAGC	AGAGGACAGC	13860
50	AGATGGCTCC	ACCTGCTGAG	GAAGGAGACG	TGTTTGTGGG	TGTTACAGGG	ATGGTGCTGC	TGGGCCCTGC	13930
	CGTGTCCTCA	CCCTGTTTTT	CTGGATTGTA	TGTTGAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGCTCCAG	14000
	TGCTCCAGAG	CCCTACCGTG	GCAGGTAGAA	GAAGTCCCGA	TTTACCCCC	TCCCCACAAA	CTCCCCAAGC	14070
	ATGTAAGACT	TCCGGCCATG	CAGACAAGGA	GGGTGACCTT	CTTGGGGGTC	TTTTTTTTCT	TTTTTTCTTT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTGTAC	AGTGAGAAT	14210
55	TGCTAACTCG	GCGGTGTTTA	CAGCAGGTGG	CTTGAAATGC	TGCGTCTTGC	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACGGC	AGCTGCCCTCA	CACCTGCTGC	GGCTCAGGTG	GACCACGCGG	AGTCAGATAA	GCGTCATGCA	14350
	ACCCAGTTTT	GCTTTTTGTG	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTTCTCTGAT	CAGGACTCTG	14420
	CCTGTCAATTG	CTGTTCTCTG	ACTTCAGATG	AGGTCACAAT	CTGCCCTGCG	CTTATGCAAG	GAGTGAGGCG	14490
	TGGTCCCCCG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTCACGTGTA	14560
60	GGGTGAGTGA	GGCGCGGCCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	ATCCCCGGGT	GTCCTGTCA	CGTGTAGGCT	GAGTGAGGCG	14700
	TGGTCCCCCG	GTGTCCCTGT	CCCGTGCAGG	GTGAGTGAGG	CAGTGTCCCC	GGGTGTCCCT	GTCACGTGCA	14770
	GGGTGAGTGA	GGCGCGGCTC	CCGGGTGTCC	CTCTCAGGTG	TAGGGTGAGT	GAGGCGCGCG	CCGAGGGTGT	14840
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCAAC	GTCCTGGGT	GTCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCCCG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGCGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCTGTCC	CTGGGTGTCC	CTGTCTCGTG	TAGGGTGAGT	GAGGCTCTGT	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGCTCC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GGCGCGGTTG	15120
	CCCATTCGCT	GGGTAGATGG	TGCAGGCGCA	GTGCTGGTCC	CCAAGCCTAT	CTTTTCTGAT	GCTCGGCTCT	15190
	TCTTGGTCAC	CTCTCCGTTT	CATTTTGCTA	CGGGGACACG	GGAGTGCAGG	CTCTCGCTC	CCCGGTGCCA	15260
70	GGCACTGCAG	CCACAGCTTC	AGGTCCGCTT	GCCTCTGTTG	GGCCTGGCTT	GCTCACCACG	TGCCCGCCAC	15330
	ATGCCATGCTG	CCAATACTCC	TCTCCAGCTG	TGTCTCATGC	CGAGGCTGGA	CTCTGGGCTG	CTGTGTCTGT	15400
	CTGCCACGTG	TTGCTGGAGA	CATCCAGAA	AGGGTTCTCT	GTGCCCTGAA	GGAAAGCAAG	TACCCCCAG	15470
	CCCTCACTTT	GTCCTGTTTT	CTCCCAAGCT	GCCCTCTGCG	TTGGCCCCCT	TGGGTGGGTG	GCACGCTTGT	15540
	TACCTTTATT	CTGGGCACCT	GCCGCTCATT	GCTTAGGCTG	GGCTCTGCTT	CCAGTCCGCC	CCTCACATGG	15610
75	ATTACGCTCC	AGCCACAGGT	TGGAGTGTCT	CTGTCTGTCT	CCTGCTCTGA	GACCCACGTG	GAGGGCCGCT	15680
	GTCTCCGCGA	GCCTTCGTCA	GACTTCCCTC	TTGGGTCTTA	GTTTGTGAAT	TCACTGATTT	ACCTCTGACG	15750
	TTCTATCTCT	TCCATTGTAT	GCTTTTCTCT	GTTTTATTCT	TTTATTCTCT	TTTAGTCTTC	TTAGTTTAGT	15820
	CATGCCTTTC	CCTCTAAGTG	CTGCCTTACC	TGCACCTGT	GTTTGTGATG	GAAGTAATCT	CAACATCAGC	15890

	CACTTTCAAG	TGTTCTTAAA	ATACTTCAAA	GTGTTAATAC	TTCTTTTAA	TATTCTTATT	CTGTGATTTT	15960
	TTTCTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTTTG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCTGTGA	TTTCTTTGAG	CAGTGAGTTA	TTTGAACACT	GTTTATGTTC	AAGATATGTA	GAGTATCAAG	16100
5	ATACGTAGAG	TATTTTAAAGT	TATCATTTTA	TTATTGATTT	CTAACTCAGT	TGTGTAGTGG	TCTGTATAAT	16170
	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGTGATCT	AGTGTGTGCA	TGGTTTCCAG	AACTGTCCAT	16240
	TGTAATTTTG	ACATCCTGTG	AATAGTGGGC	ATGCATGTTC	ACTATATCCA	GCTTATTAAG	GTCCAGTGCA	16310
	AAGCTTCTGT	CTCCTTCTAG	ATGCATGAAA	TTCCAAGAG	GAGGCCATAG	TCCCTCACCT	GGGGATGGG	16380
	CTGTTCATT	TCTTCTCGTT	TGGTAGCATT	TATGTGAGGC	ATTGTTAGGT	GCATGCACGT	GGTAGAATT	16450
10	TTATCTTCCT	GATGAGTGAA	TCTTTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCTCTG	TGTGTCTGTT	TTCTGCCTTT	16590
	AATTTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTCCGCCA	GGGTGAGTGC	16660
	AGTGGTGTGA	TCACAGGTCA	GTGTAACCTT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCGCTA	CACCTGGCTA	ATTTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
15	TGCTGTGTG	CCCAGGCTGG	TCTCAAACTC	TTGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
	CTGAATTACA	GGCATGAGCC	ACCATGTCTG	GCCTAATTTT	CAACACTTTT	ATATTTCTAT	AGTGTGGGTA	16940
	TTCCCTGTTA	ACAGCATGTA	GGTGAATTC	CAATCCAGTC	TGACAGTCGT	TGTTTAACTG	GATAAGCTGA	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGCATGT	17080
	CCTCGTTCCC	TTGTTTCTCA	CCACCTCTTG	GGTTGCCATG	TGCGTTTCTT	GCCGAGTGTG	TGTTGATCCT	17150
20	CTCGTGTGCT	CCTGGTCACT	GGGCATTGCG	TTTTATTCTT	CTTGTCTTAG	TGTTACCCCTG	TGATCTTTTT	17220
	ATTGTCGTTG	TTTGCTTTTG	TTTATTGAGA	CAGTCTCACT	CTGTCAACCA	GGCTGGAGTG	TAATGGCACA	17290
	ATCTCGGCTC	ACTGCAACCT	CTGCCCTCTC	GGTTCAAGCA	GTTCTCATTC	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	CGCCCAACCC	CACGCTGGGC	TAATTTTGTG	ATTTTATAGT	GAGATAGGCT	TTCAACATGT	17430
	TGGCCAGGCT	GGTCTCAAAC	TCTTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACAG	TGCTGGGATT	17500
25	ACAGGTGCAA	GCCACCGTGC	CCGGCATACC	TGTATCTTTT	AAATGAAAGT	CTGAAACATT	GCTACCTTTG	17570
	TCCTGAGCAA	TAAGACCTTT	AGTGTATTTT	AGCTCTGGCC	ACCCCCAGC	CTGTGTGCTG	TTTCCCTGCG	17640
	TGACTTAGTT	CTATCTCAGG	CATCTTGACA	CCCCCAACAG	CTAAGCATT	TTAATATTGT	TTTCCGTGTT	17710
	GAGTGTCTCT	GTAGCTTTGC	CCCCGCCCTG	CTTTTCTCTC	TTTGTTCCTC	GTCTGTCTCT	TGCTCTCAGG	17780
30	CCGCGCTCTG	GGGTCCCTCT	CCTTGTCTCT	TGCGTGTTTC	TTCTGTCTTG	TTATTGCTGG	TAAACCCAGG	17850
	CTTTACCTGT	GCTGGCTTCC	ATGGCATCTA	GCGACGTCCG	GGGACCTCTG	CTTATGATGC	ACAGATGAAG	17920
	ATGTGGAGAC	TCACGAGGAG	GGCGGTCTAT	TGCGCCGTG	AGTGTCTGGA	GCACCCAGTG	GCCACGGTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCCGCT	CGCCCTGGGT	TCAGCTGGA	AAACCCAGG	CATGTCCGGG	18060
	TCGTGGTGGT	CCGCGGTGTC	GAGTTTGAAG	TCGCGCAAC	CTGCGGTGTG	GCGCCAGCTC	TGACGGTGTCT	18130
	GCCTGGCGGG	GGAGTGTCTG	CTTCTCTCTC	TCGTGTTGGG	AACCAAGGAC	AAGGATGAGG	CTCCAGCCCG	18200
35	TTGTGCGCCA	ACAGGAGCAT	GACGTGAGCC	ATGTGGATAA	TTTTTAAAT	TCTAGGCTGG	GCGCGGTGGC	18270
	TCACGCCCTG	AATCCACGCA	CTTTGGGAGG	CCAAGGCGGG	TGGATCACGA	GGTCAGGAGG	TGCTGACCAT	18340
	CCTGGCCAAC	ATGATGAAAC	CCCATCTGTA	CTAAAAACAC	AAAAATTAGC	TGGCGTGGT	GGCGGGTGCC	18410
	TGTAATCCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	GTTCAGCTGA	18480
40	GCCGACATTG	CACCACTGCA	TCCAGCGCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAATA	18550
	AAAAAAAATA	AATTTCTAGT	GCCACATTAA	AAAAGTAAAA	AAGAAAAGGT	GAAATTAATG	TAATAATAGA	18620
	TTTTACTGAA	GCCACAGCAT	TCCACACCTC	ATCATTTTAG	GGTGTATTG	GTGGGAGGAT	CACCTACAGG	18690
	ACATTGACAA	TTTTTTGAGC	TTTGTCTGCG	GGATCCCGTG	TGTAGTCTCC	GTGCGTGGCC	ATCTCGGCCT	18760
	GGACCTCTG	GGCTTCCCAT	GGCCATGGCT	TTGTACCAG	ATGGTGCAGG	TCCGGGATGA	GGTCGCCAGG	18830
45	CCCTCAGTGA	GCTGGATGTG	CAGTGTCCGG	ATGGTGCACG	TCTGGGATGA	GGTCGCCAGG	CCCTGCTGTG	18900
	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTCAGGGGTG	AGGTCTCCAG	GCCCTCGGTG	AGCTGGAGGT	18970
	ATGGAGTCCG	GATGATGCAG	GTCGGGGGTG	AGGTGCGCCAG	GCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GCCCTCGGTG	AAGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	GGTCCGGGGT	GAGGTGCGCA	GGCCCTGCTG	TGAGCTGGAT	GTGTGGTGTG	TGGATGGTGC	AGGTCTGGGG	19180
	TGAGGTCAAC	AGGCCCTGCG	GTGAGCTGGG	TGTCCGGTGT	CTGGATGGTG	CAGGTCTGGA	GTGAGGTCCG	19250
50	CAGACGGTGC	CAGACCATGC	GGTGAGCTGG	ATATGCGGTG	TCCGGATGGT	GCAGGTCTGG	GGTGAGGTTC	19320
	CCAGGCCCTG	CTGTGAGTGT	GATGTGGGTG	GTCGGGATGC	TGCAGGTCCG	GTGTGAGGTC	ACCAGGCCCT	19390
	GCTGTGAGCT	GGATGTGTGG	TGTCTGGATG	GTGCAAGTCT	GGGTGGAAGG	TCCGCAAGGC	CCTGCTTGTG	19460
	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GCTGGAGTGT	AGGTGCGCCAG	GCCCTCGGTG	AGCTGGATCT	19530
	GCAGTGTCCA	GATGGTGCAG	GTCCGGGGTG	AGGTGCGCCAG	ACCTTCCGGT	GAGCTGGATG	TCCGGTGTCT	19600
55	GGATGGTGCA	GGTCTGGAGT	GAGGTGCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGGTGCC	19670
	GGTCCGGGGT	GAGGTGCGCA	GACCCTGCTG	TGAGCTGGAT	GTGCGGTGTC	TGGATGGTAC	AGGTCTGGAG	19740
	TGAGGTGCGC	AGACCTGCTC	GTGAGCTGGA	TATGCGGTGT	CCGGATGGTG	CAGGTGAGGG	GTGAGGTCTC	19810
	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTGCGGG	GTGAGGTGCG	CAGGCCCTCG	19880
	TGTGAACCTG	ATGTGCGGGG	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTGC	CCAGGCCCTG	GGTGAGCTGG	19950
60	AGGTATGGAG	TCCGGATGAT	GCAGGTCCGG	GGTGAGGTGC	CCAGGCCCTG	CTGTGAGCTG	GATGTGCGGG	20020
	GCTCTGGATG	TGCAGGTCTG	GGGTGTGCTG	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20090
	TGCAGGTCCG	GGGTGAGGTT	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTCAGTCCG	20160
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGGTC	GGGTGAGGTC	20230
	CACCAAGGCC	TGCGGTGAGC	TGGTTGTGCG	GTGTCCGGTT	GCTGCAGGTC	CGGGGTGAGT	TGCGCAAGGC	20300
65	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGGT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	CTAGGGCCCT	20370
	TGGTGGGCTG	GATGTGCGGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGTTGAGCTG	20440
	GATGTGCGGT	GTCTGCATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATCTGTGGT	20510
	GTCCGGATGG	TGCAGGTCCG	GCTGTGAGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGCTGTGATG	20580
	GTGCAAGTCC	GGGGTGAGGT	AGCCAAGGCC	TTCGGTGAGC	TGGATGTGGG	GTGTCCGGAT	GGTGCAAGTC	20650
70	CGGGGTGAGG	TCGCCAGGCC	CTGCCGTTAG	CTGGATATGC	GGTGTCCGGA	TGGTGCAGGT	CCGGGGTGAG	20720
	GTCACCAAGG	CCTGCGGTTA	GCTGGATGTG	CGGTGTCTGG	ATGGTGCAGG	TCCGGGGTGA	GGTCCGAGG	20790
	CCCTGCTGTG	AGCTGGATGT	GCTGTATCCG	CATGGTGCAG	GTCCGGGGTG	AGGTGCGCAG	CCCTGCGAGT	20860
	GAGCTGGATG	TGCTGTATCC	GGATGGTGCA	GGTCTGGCGT	GAGGTGCGCA	GGCCCTGCGG	TTAGCTGGAT	20930
	ATCGGGTGTG	GGATGGTGCA	GGTCCGGGGT	GAGGTACCA	GGCCCTGCGG	TTAGCTGGAT	GTGCGGGTGT	21000
75	CGGATGGTGT	AGGTCTGGGG	TGAGGTGCGG	AGGCCCTGCT	GTGAGCTGGA	TGTGCTGTAT	CCGATGGTGT	21070
	CAGGTCCGGG	GTGAGGTGCG	CAGGCCCTGC	GGTGTGAGCT	ATGTGCTGTA	TCCGGATGGT	GCAGGTCTGG	21140
	CGTGAGGTGC	CCAGGCCCTG	CGGTGAGCTG	GATGTGCAAT	GTACGGATGG	TGCAGGTCCG	GGGTGAGGTC	21210
	GCCAGGCCCT	GCGGTGGGCT	GTATGTGTGT	TGTCTGATG	GTGCAAGTCC	GGGTGAGT	GCCAGGCCCT	21280

5	TGCGGTGAGC	TGGATGTGTG	GTGCTCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TGCGCCAGGCC	CTCGGTGAGC	21350
	TGGATATCGG	GTGTCCCGGT	GTCCGAATGG	TGCAGGTCCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	21420
	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGAGCTG	GATGTGCCGT	21490
	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCTTCTCG	21560
	TTTAAAGGGT	TGGCTGTGTT	CGGGCCGAG	AGCACCGTCT	GCGTGAGGAG	ATCCTGGCCA	AGTTCCTGCA	21630
	CTGGCTGATG	AGTGTGTACG	TCGTGAGCT	GCTCAGGTCT	TTCTTTTATG	TCACGGAGAC	CACGTTTCAA	21700
	AAGAACAGGC	TCTTTTCTA	CCGGAAGAGT	GTCTGGAGCA	AGTTGCAAA	CATTGGAATC	AGGTACTGTA	21770
	TCCCCAGGCC	AGGCCTCTG	TTCTCGAAGT	CCTGGAACAC	CAGCCCGGCC	TCAGCATGCG	CCTGTCTCCA	21840
	CTTGCCCTGTG	CTTCCCTGGC	TGTGCACTG	TGGGCTGGGA	GCCAGGGGCC	CCGTGACAGG	CCTGGTCCAA	21910
10	GTGGATTCTG	TGCAAGGCTC	TGACTGCTG	GAGCTCACGT	TCTCTTACTT	GTAAATCAG	GAGTTGTGTC	21980
	CAAGTGGTCT	CTAGGGTTTG	TAAAGCAGAA	GGGATTAAAA	TTAGATGGAA	ACACTACCAC	TAGCCTCTCT	22050
	GCCTTTCCCT	GGGATGTGGG	TCTGATTCTC	TCTCTCTTTT	TTTTTCTTTT	TTTGAGATGG	AGTCTCACTC	22120
	TGTTGCCCGG	GCTGGAGTGC	AGTGGCATAA	TCTTGCTCA	CTGCAACCTC	CACCTCTGGG	GTTTAAGCGA	22190
	TTTACCAGCC	TCAGCCTCCT	AAGTAGCTGG	GATTACAGGC	ACCTGCCACC	ACGCTCTGGT	AATTTTGTGA	22260
15	CTTTTAGGAG	AGACGGGGTT	TCACCATGTT	GGCCAGGCTG	GTCTCGAACT	CATGACCTCA	GGTGATCCAC	22330
	CCACCTTGCC	CTCCCAAAGT	GCTGGGTTTA	CAGGCTAAGC	CACCGTGCCC	AGCCCCCGAT	TCTCTTTTAA	22400
	TTCATGCTGT	TCTGTATGAA	TCTTCAATCT	ATTGGATTTA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
	GGCGACTCAC	TGCAGGGAGC	ACCTGTGAGC	GGAGCACCTG	GGGATAGGAG	AGTTCCACCA	TGAGCTAACT	22540
20	TCTAGGTGGC	TGCATTGGAA	TGGCTGTGAG	ATTTTGTCTG	CAATGTTCCG	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTG	AAGCTGGATT	TGCATCAGTG	AGGACCGGGA	GCGCTGGTCT	GGGAGATGCC	AGGCTGGCTG	22680
	AGCCCAAGGC	ATGGTATTAG	CTTCTCCGTG	TCCCGCCAG	GCTGACTGTG	GAGGGCTTTA	GTCAAGAGAT	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCACACTCG	AGTCCCTGGG	GGGCCCTGTG	ACACCCCATG	CCCCAAATCA	22820
	GGATGTCTGC	AGAGGGAGCT	GGCAGCAGAC	CTCGTCAGAG	GTAACACAGC	CTCTGGGCTG	GGGACCCCGA	22890
25	CGTGGTCTG	GGGCCATTTC	CTTGATCTG	GGGGAGGGTC	AGGGCTTTCC	CTGTGGGAAC	AAGTTAATAC	22960
	ACRAATGCC	TTACTTAGAC	TTTACACGTA	TTTAATGGTG	TGCGACCCAA	CATGGTCATT	TGACCATGAT	23030
	TTTGGAAAGA	ATTTAATTGG	GGTGACCGGA	AGGAGCAGAC	AGACGTGGTG	GTCCCAAGA	TGCTCTCTGT	23100
	CACACTATGG	ACTGTGTGTC	TGCCCTGGGG	CGCTTGGAGG	CCCCCTCTCC	CTGGACAGGG	TACCGTGCCT	23170
	TTTCTACTCT	GCTGGGCTG	CGGCTCTGGG	TCAGGCGACC	AGCTCCGGAG	CACCCCGGCC	CCCATGTGTC	23240
30	ACCGAGTGCC	AGGCTGTGAG	CCACAGATGC	CCAGGTCCAG	GTGTGGCCGC	TCCAGCCCCC	GTGCCCCCAT	23310
	GGGTGGTTTT	GGGGGAAAAG	GCCAAGGGCA	GAGGTGTGAG	GAGACTGGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGCC	TGAGCTGCCC	TGAGCAGCCT	CTCCCGCCCT	CTCCATCTGA	AGGGATGTGG	CTCTTTCTAC	23450
	CTGGGGGTGC	TGCCCTGGGG	CAGCCTTGGG	CTACCCAGT	GGCTGTACCA	GAGGGACAGG	CATCCTGTGT	23520
	GGAGGGGCAT	GGGTTACAGT	GGCCCCAGAT	GCAGCCTGGG	ACCAGGCTCC	CTGGTGCTGA	TGGTGGGACA	23590
35	GTCAACCTGG	GGGTTGACCG	CCGGAAGTGG	CGTCCCCAGG	GTTGACTATA	GGACCAAGTG	TCCAGGTGCC	23660
	CTGCAAGTAG	AGGGGCTCTC	AGAGGCGTCT	GGCTGGCATG	GGTGGACGTC	GCCCCGGGCA	GTGCTTTCAG	23730
	CGTGTGCTGC	CGTGGGTGCC	CTGAGCCCTC	ACTGAGTCCG	TGGGGGCTTG	TGGCTTCCCG	TGAGCTTCCC	23800
	CCTAGTCTGT	TGCTGTGGCT	AGCAAGCCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	GAGAGGCTTG	23870
	CAGCTGCGGG	AGCTGTGCGA	AGCAGAGGTC	AGGCAGCATC	GGGAAGCCAG	GCCCCGCCCT	CTGACGTCCA	23940
40	GACTCCGCTT	CATCCCCAAG	CCTGACGGGC	TGCGGCCGAT	TGTGAACATG	GACTACGTCG	TGGGAGCCAG	24010
	AACGTTCGCC	AGAGAAAAGA	GGGTGGCTGT	GCTTTGGTTT	AACTTCCTTT	TAAACAGAAA	GTGCTTTTGA	24080
	GCCCACATCT	TGGTATCAGC	TTAGATGAAG	GGCCCGGAGG	AGGGGCCACG	GGACACAGCC	AGGGCCATGG	24150
	CACGGCGCCA	ACCCATTGTT	CGCGACAGTG	AGGTGGCCGA	GGTGCCGGTG	CCTCCAGAAA	AGGAGGCTTG	24220
	GGGTGTAGGG	GGAGCTCCTG	GGGCAGGGAC	AGGCTCTGAG	GACCACAAGA	AGCAGCCGGG	CCAGGGCTGT	24290
45	GATGCGACAC	GGCCCGAGGT	CTTGGATCCG	TGTCCTGCTG	TGGTGCCGAG	CCTCCGTGCG	CTTCCGCTTA	24360
	CGGGGCCCGG	GGACCAAGCC	ACGACTGCCA	GGAGCCCAAC	GGGCTCTGAG	GATCCTGGAC	CTTCCCCCAC	24430
	GGCTCTGCA	CCCCACCCCT	GTGGCTGCGG	TGGCTGCGGT	GACCCCGTCA	TCTGAGGAGA	GTGTGGGGTG	24500
	AGGTGGAGAC	AGGTGTGGCA	TGAGGATCCC	GTGTGCAACA	CACATGCGGC	CAGGAACCCG	TTTCAAAACAG	24570
	GGTCTGAGGA	AGCTGGGAGG	GGTTCTAGGT	CCCGGGTCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
50	TCTCCCTGG	GTCCCTATGG	TGGGGTGGGC	ACTTGGCCGG	ATCCACTTTC	CTGACTGTCT	CCCATGTGCT	24710
	CCCCGCCAGG	CGGAGCTGCT	CACCTCGAGG	GTGAAGGCAC	TGTTACAGCT	GCTCAACTAC	GAGCGGGCGC	24780
	GGCGCCCGGG	CCTCCTGGGC	GCCTCTGTGC	TGGGCCCTGA	CGATATCCAC	AGGGCCTGGC	GCACCTTCTG	24850
	GCTGTGTGTG	CGGGCCCGAG	ACCCGCCCGC	TGAGCTGTAC	TTTGTCAAGG	TGGGTGCCGG	GGACCCCCGT	24920
	GAGCAGCCCT	GCTGGACCTT	GGGAGTGGCT	GCCTGATTGG	CACCTCATGT	TGGGTGGAGG	AGGTACTCTT	24990
55	GGGTGGGCGG	CAGGGAGTGC	AGGTGACCCT	GTCACTGTTG	AGGACACACC	TGGCACCTAG	GGTGGAGGCC	25060
	TTACAGCTTT	CCTGAGCAGC	ATGGGGCCGA	CTGTGCAACC	TGACTGCCCC	GGCTCCTATT	CCCAAGGAGT	25130
	GTCCCACTGG	ATTCCAGTTT	CCGTGAGAGA	AGGAACCGCA	ACGGCTCAGC	CACCAAGGCC	CGGTGCCCTT	25200
	CACCCAGCTC	CTGAGCCAGG	GGTCTCCTGT	CCTGAGGCTC	AGAGAGGGGA	CACAGCCCGC	CCTGCCCTTG	25270
	GGGTCTGGAG	TGGTGGGGGT	CAGAGAGAGA	GTGGGGGACA	CCGCCAGGCC	AGGCCCTGAG	GGCAGAGGTT	25340
60	ATGCTGAGT	TTCTGCGTGG	CCACTGTGAG	TCTCCTCGCC	TCCACTCACA	CAGGTGGATG	TGACGGGGCC	25410
	GTACGACACC	ATCCCCCAGG	ACAGGCTCAC	GGAGGTCATC	GCCAGCATCA	TCAAACCCCA	GAACACGTAC	25480
	TGCGTGCCTG	GGTATGCCGT	GGTCCAGAAG	GCCGCCCATG	GGCAGCTCCG	CAAGGCCTTC	AAGAGCCACG	25550
	TAAGGTTTCA	GTGTGATAGT	CGTGTCCAGG	ATGTGTGCTC	CTGGGATATG	AATGTGCTTA	GAATGCAGTC	25620
	GTGTCTGTGA	TGCGTTTCTG	TGGTGGAGGT	ACTTCCATGA	TTTACACATC	TGTGATATGC	GTGTGTGGCA	25690
65	CGTGTGTGTC	GTGGTGCATG	TATCTGTGGC	GTGCATATTT	GTGGTGTGTG	TGTGTGTGGC	ACGTGTGTGT	25760
	CCATGGTGTG	TGTGCTGTG	GTGTGCTGTC	GTGTGTGCTC	GTGACACGTG	CATGTTCTATG	CTGTGTGCTG	25830
	CATGCTGTG	ATGTGCCTAT	TTGTGGTGTG	TGTGTGCATG	TGTCCGTGAC	ATATGCGTGT	CTATGGCATG	25900
	GGTGTGTGTG	GCCCCCTTGG	CTTACTCCTT	CCTCCTCCAG	GCATGGTCCG	CACCATTTGT	CTCAGCTCTC	25970
	CGGGTGTCTG	TTTGGGGAGC	TCCACATTCA	GGGTCTCTAC	TTCTAGCATG	GGTGCCCTTG	TCTGTGCACA	26040
70	GGGCTGGGCC	TTGGAGACTG	TAAAGCCAGT	TTGAGAGGAG	AGTAGGGATG	CTGGTGGTAC	CTTCTGGGAC	26110
	CCCTGGCACC	CCCAGGACCC	CAGTCTGGCC	TATGCCGGCT	CCATGAGATA	TAGGAAGGCT	GATTCAGGCC	26180
	TCGCTTCCCG	GGACACACTC	CTCCAGAGGC	GGCCGGGGGC	CTTGGGGCTC	GGCAGGGGTT	AAAGGGGGCC	26250
	TGGGCTTGGG	TTCCCAACCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTCAAAGTCG	TGCCAGGGCC	26320
	GGGTGCGAGG	GTGAAGAAGT	ATCCCTGGAG	CTTCGGTCTG	GGGAGAGGCA	CATGTGGAAA	CCCAACAGGA	26390
75	CCTCTTTCTC	TGACTTCTTG	AGCT					26414

## Contig 2:

	TGTGGGATTG	GTTTTCATGT	GTGGGATAGG	TGGGGATCTG	TGGGATTGGT	TTTTATGAGT	GGGGTAACAC	70
	AGAGTTC AAG	GCGAGCTTTC	TTCTCTGAGT	GGGTCTGCAG	GTGCTCCAAC	AGCTTTATTG	AGGAGACCAT	140
5	ATCTTCCTTT	GAACATATGT	CGGGTTTATA	GTAAGTCAGG	GGTGTGGAGG	CCTCCCTTGG	GCTCCCTGTT	210
	CTGTTTCTTC	CACCTCTGGG	TCGTGTGGTG	CCTGCTGTGG	TGTGTGGCCG	GTGGGCAGGG	CTTCCAGGCC	280
	TCCTTGTGTT	CATTGGCCTG	GATGTGGCCC	TGGCTACGCT	CCGTCTCTGG	AATCCCTCTG	CGAGTTGGAG	350
	GCTTTCCTTC	TTTCTTTTTT	TTTCTCTTTT	TTTTTTTTTT	TGATAACAGA	GTCTCGCTCT	TTTTTGCCCA	420
	GGCTGGAGTG	GTTTGGCGTG	ATCTTGGCTC	ACTGCAACCT	GTGCTTCCTG	AGTTCAAGCA	ATTCTCTTGC	490
10	CTCAGCCTCC	CAAGTAGCTG	GAATTATAGG	CGCCACACAC	CATGCTGACT	AATTTTTGTA	ATTTTAGTAG	560
	AGACGAGGTT	TCTCCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	GGTGATCCTC	CCACCTCGGC	630
	CTCCCAAAGT	GCTGGGATGA	CAGGTGTGAA	CGCCCGCGCC	CGGCCGAGAC	TCGCTTCTTG	CAGCTTCCGT	700
	GAGATCTGCA	GCGATAGCTG	CCTGCAGCCT	TGGTGCTGAC	AACCTCCGTT	TTCCTTCTCC	AGGTCTCGCT	770
	AGGGGTCTTT	CCATTTTCATG	ACTCTCTTCA	CAGAAGAGTT	TCACGTGTGC	TGATTTCCCG	GCTGTTTCTT	840
15	CGCTAATTGG	TGTCTGCTGT	TTATCGATGG	CCTCCTTCCA	TTTCCTTTAG	GCTTTGTTTA	TGTTGTTT	910
	TCCGGCTCCT	TGAAGGAAAA	GTTTCGATTA	TGGATGTTTG	AACCTTCTTT	TCTAAACAAG	CATCTGAAGT	980
	TGCCGTTTTT	CCTCTAAAGC	AGGGATCCCG	AGGCCCTTGG	CTGTGGAGTG	GCACCGGTCT	GGGGCTGTT	1050
	AGGAACCCGG	GCGACACCGG	GAGGCTAGGT	GGGGTGTGGG	GAGCCAGCGT	TCCCGCCTGA	GCCCGGCCCC	1120
	TCTCAGATCA	CGAGTGGCAT	GCGGTGCTCA	GAGGCGCACA	CACCCTACTG	AGAACTGTGC	GTGAGAGGGG	1190
20	TCTAGATTCT	GTGCTCCTTA	TGGGAATCTA	ATGCCCTGATG	ATCTGAGGTG	GAACCGTTTG	CTCCCAAAAC	1260
	CATCCCTTTC	CCCCTGCTG	TCCTGTGGAA	AAATCGTCTT	CCACGAAACC	AGTCCCTGGT	ACCACAATGG	1330
	TTGGGGACCC	TGTGCTAAAG	ACCTGCTTCA	GCAGCCTCTC	GTGAGTGTG	ATATATTGGC	TTTTCTGTGT	1400
	TGAGTCCAGA	ATAATTACGG	ATTTCTGTGA	TGCTTTCGCG	CGACCTCAGA	CCCATGGGCT	ATTTGTGGGC	1470
	GTGTTGCCTG	CTCCTGGGTT	GGGAAGGGTG	CAGGCCCCAT	GTACCTTCTT	GTTACTGCCT	TCCAGGTTGG	1540
25	TTCTCAGGGT	TGAATCGTAC	TCGATGTGGT	TTTAGCCACAC	GGCCCTGCGG	CCAGCTCCTG	GGGGCTGGGG	1610
	AACATGCTGA	AGCAGAGAGT	CACCGTGGCG	GTCTTTTGAT	GCCTCACAAG	CTCAGGCGCT	CCTGTGTCCG	1680
	TGTTAGTGTG	TGTCACGTGC	CTGCTCACAT	CTGTCTTGG	GGACGCGAGG	GCTTAGCAGG	TCCCGTAGTA	1750
	AATGACAAGC	GTCTGCGGGG	AGTCTGCAGA	ATAGGAGGTG	GGGGTGCCGG	TCTCTCTCCC	CGGTCTTCAG	1820
	ACTCTTCTCC	TGCCCTGTCT	GTGGCTGCAC	CTGCATCCCT	GCAATCCCTC	CAGCACTGGG	CTGGAGAGGC	1890
30	CCGGGAGCTC	GAGTGCCACT	TGTGCCACGT	GACTGTGGAT	GGCAGTCGGT	CACGGGGGTC	TGATGTGTGG	1960
	TGACTGTGGA	TGGCGGTTGG	TCACAGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2030
	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2100
	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	GACTCTGGAT	GGCGGTCTGT	GGGTCTGATG	2170
	TGGTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2240
35	TGGTGACTGT	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	2310
	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2380
	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2450
	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTGC	TGGGGTCTGA	2520
	GGTCTGATGT	GTGGTGACTG	TGGATCGTGA	TGGTCTGATG	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	2590
40	GTGCTGGGGT	CTGATGTGTG	TGGATCGTGA	TGGTCTGATG	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	2660
	GATGGCGATC	GGTCACAGGG	GTCTGATGTG	TGGTGACTGT	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2730
	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	2800
	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	2870
45	GACTGTGGAT	GGCGGTCTGT	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	2940
	GTGGTGACTG	TGGATGGCAG	TCGTGGGGTC	TGATGTGTGG	TGACTGTGGA	TGGCGGTCTG	GGGTCTGATG	3010
	TGTGGTGACT	TGGATGGCAG	TCGTGGGGTC	TGATGTGTGG	TGACTGTGGA	TGGCGGTCTG	GGGTCTGATG	3080
	GTGTTGGTGG	ACTGTGGATG	GGCGGTCTGT	GGTCTGATGT	GGTGACTGTG	GATGGCGGTG	GTGGGGTCTG	3150
	ATGTGTGGTG	ACTGTGGATG	GTGATCGGTC	ACAGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGT	3220
50	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3290
	GGGTCTGATG	TGTGGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3360
	GTACACAGGG	TCTGATGTGT	GGTGACTGTG	GATGGCGGTG	GTGGGGTCTG	ATGTGTGGTG	ACTGTGGATG	3430
	CGGGTCGTGG	GGTCTGATGT	TGGTGACTGT	TGGATGGCGG	TCGTGGGGTC	TGATGTGTGG	TGACTGTGGA	3500
	TGGCGGTCTG	GGGGTCTGAT	GTGGTGACTG	TGGATGGTGA	TCGGTCAACG	GGGTCTGATG	TGTGGTAGCT	3570
55	GCAGGTGGAG	TCCCAGGTGT	GTCTGTAGCT	ACTTTGCGTC	CTCGGCCCCC	CGGCCCCCGT	TTCCCAAAAC	3640
	GAAGCTTCCC	AGGCGCTCTC	TGGGCTTCAT	CCGCCCATCG	GGCTTGGCCG	CAGGTCCACA	CGTCTGATC	3710
	GGAAGAAACA	AGTCCCCAGC	TCTGGCCGGG	GCAGGCCACA	TTTGTGGCTC	ATGCCCTCTC	CTCTGCCGGC	3780
	AGGTCTCTAC	CTTGACAGAC	CTCCAGCCGT	ACATGCGACA	GTTCGTGGCT	CACCTGCAGG	AGACCAGCCC	3850
	GCTGAGGGAT	GCCCTCGTCA	TCGAGCAGGT	CTGGGCACTG	CCCTGCAGGG	TGGGCGACGG	ACTCCACGCA	3920
60	GTGGGTCCTC	CCCTGGGCAA	TCACTGGGCT	CATGACCCGA	CAGACTGTTG	GCCCTGGGGG	GCAAGTGGGG	3990
	GAATGAGCTG	TGATGGGGGG	ATGATGAGCT	GTGTGCTTGG	GCGAAATCTG	AGCTGGGGCC	TGCCAGGCTG	4060
	CGACAGCTGC	TGCATTACAG	CACCTGCTCA	CGTTTGACTG	CGCGGCTCTC	CTCCAGTTCC	GCAGTGGCTT	4130
	TGTTTATGAT	TTGCTAAATG	TCTTCTCTGC	CAGTTTTGAT	CTTGAGGCCA	AAGGAAAGGT	GTCCCTCTCC	4200
	TTTAGGAGGG	CAGGCCATGT	TTGAGCCGTG	TCTGCCCCAG	CTGGCCCCCT	AGTGCTGGGT	CTGAGGCCAA	4270
65	AGGAAACGTG	TCCCTCTTCT	TAGGAGGACG	GGCCGTGTTT	GAGCCACGCC	CCGCTGAGCG	GGCCTCTCAG	4340
	TGCTGGGTCT	GTCCACGTGG	CCCTGTGGCC	CTTTGCAGAT	TGGGTCTGTC	CACGTGGCCC	TGTGGCTCTT	4410
	TGCAGATGCC	TGTTAGCACT	TGCTCGGCTC	TAGGGGACAG	TCGTGTCCAC	CGCATGAGGC	TCAGAGACCT	4480
	CTGGGCGAAT	TTCTTGGGCT	CCCAGGGTGG	GGGTGGAGGT	GGCCTGGGCT	GCTGGGACCC	AGACCTCTGT	4550
	CCCGGCAGCT	GGGACGCAAC	TCTTGGATCA	CATATGCCAT	CCGGGCCACG	GTGGGCTGTG	TGGGTGTGAG	4620
70	CCCAGCTGGA	CCCACAGGTG	GGCCAGAGGA	GACGTTCTGT	GTACACACTC	CTGCCCTAAG	CCATGTGTGT	4690
	CTGCAGAGAC	TGGCCCCGGC	CAGCCACAGA	TGGCCCTGCA	TTCCAGCCCA	GCCCCGCACT	TCATCACAAA	4760
	CACGTACCCC	AAAAGGGGAG	GAGGGTCTTG	GCCACGTGGT	CCTGCTGTCT	TCAGCACCCA	CCGGCTCACT	4830
	CCCATGTGTC	TCCCGTCTGC	TTTCCGACAG	CTCCTCCCTG	AATGAGGCCA	GCAGTGGCCT	CTTGCAGCTC	4900
	TTCTACGCTT	TCAATGTGCA	CCACGCCGTG	CGCATCAGGG	GCAAGTGAGT	CAGGTGGCCA	GGTGGCAATT	4970
	CCCTGCGGGT	GGCTGGGCGG	GCTGGCAGGG	CTTCTGCTCA	CCTCTCTCCT	GCCCTTCTCC	CAGTGNCTTT	5040

	CTGCCCCGGG	CCACCAGAGT	CTCCTTTTCT	GGCCCCCGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCCG	5110
	GGCCCCGGAA	ACATGGCTCG	GCTTGGCGCA	CGCCGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGGC	5180
	AAGCGGGGTG	TGGAGTTGCT	CTCGCTGGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGGGCC	5250
5	GAGCGTTTGA	GCCTGCAGCT	TGTGAGTCC	AAGTTACTAC	TGACGCTGGA	CACCCGGCTC	TCACACGCTT	5320
	GATCTCTCT	CTCCCGATAC	AAAAGGATTT	TATCCGATTC	TCATTCTCTGT	CCCTGTCTGT	TGACCCCGCC	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTTCCTATC	TGGAAAGTGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCTCTCGG	GGGGCCTGTG	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTACAC	CAGCCACTGG	5530
	GTGAGCCACA	CTACCGGTGG	TAGAGCCACA	GTGCCGTGGT	CCACATCAGC	TCTCTGGAT	TTTAAAGTAA	5600
10	ACCACACACC	TCCCCGCGAG	CATCTGCTGT	CGACCTGTGT	TGTGCCCTGG	GAGAGTGGTA	GCACGGAGGA	5670
	AATTCGTGCA	CACTCAAGGT	CATCAGCAAG	GTGATCCGCA	GTGAGGTGGA	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGTCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAAAT	5810
	ATTGCATTAT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTATTATA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAACACAAA	TTGCACATGG	CAGCAGAGTG	AATTTTGGCC	GAGGGACACG	5950
15	TGTGCACATG	TGTGTAAGCG	CCCCCAGGCC	CCACAGAATT	CGCTGACAAA	GTACCTCCCC	CAGAGAAGCC	6020
	ACCACGGGCC	TCCTTCGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCGTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAATGTGAGG	TGATGACTGC	GTCTCATGCG	CCTGACAGAC	AGGAGGTGAC	TGTGTCTGTC	6160
	CTGTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGGCT	CTGAATAAAA	6230
	ACGTCTTCAA	AACCTGTTGC	CCCAAAACT	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCACAGGGG	6300
20	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	CGCGGACTCC	TAGAGTTGGT	CGGTGTGCTT	CTGTGCAAAA	6370
	AGTGCAGTCC	TCTTGCCCAT	CACCTGTGTA	TCTGCACCAG	CAAGGAAAGC	CTCTTTTCTT	TTCTTTCTTT	6440
	TTTTTTTTTT	GAGACGGAAAC	GTCACTGTTG	TCTGCCCTGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCCAGCATT	TCTCCTGCCCT	CAGCTCCCCG	AGCAGCTGAG	ATTACAGGCA	6580
	CCCAACCCCT	CGCGCTGGCT	AATTTTGTGA	TTTTAGTAG	AGAGGGGTTT	TTGCCATGTT	CGCCAGGCTG	6650
25	GTCTGAACT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCACGCC	CAGCCGGAAA	GCCTTTTTTT	AAGGTGACCA	CCTATAGCGC	TTCCCGAAAA	TAACAGGTCT	6790
	TGTTTTTGCA	GTAGGCTGCA	AGCGTCTCTT	AGCAACAGGA	GTGGCGTCC	GTGGGCTCTG	GGGATGGCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCACCGGG	GCTATTCTGC	TCTCACTGTT	6930
30	TGTCTGAAAA	CGCACCCCTG	GCATCCTTGT	TGGAGAGATT	TCTGCTTCTC	GTGGTTCATG	CTGAAACTAG	7000
	GGGCAAGGTT	GTATCCGTTG	CGCGCCAGCG	GCTACATGTA	GGGTTCATGAG	TCTTTCACCG	TGGACAAATT	7070
	CCTTGAAAAA	AAAAAAGGA	GTCCGGTTAA	GCATTATTTC	CGGGTCAAGT	GTCTGGTTCT	GTGAATAAAC	7140
	TCTAAGATTT	AAGAAACCTT	AATGAAGAAA	AACCTTGATC	ATTGAGAGCA	AGGATGTGGT	CACACCTGTG	7210
	GCTGGATCTG	TTTCAGCCGC	CCCACTGCAAT	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGTG	CAGAGGTCTC	7280
35	ATCTGTTATG	TTTCTGAGGT	GTTCGCCGGC	TGAATGGTAG	ACGTGTCTGT	TGTGTCTATG	AGGTCTGTGT	7350
	TCTGTGTGTG	GCTCGGTTTG	AGTGATCCGA	TGTCCAGCAC	ATGCCCTGCC	CGTCTCTCAC	CTGTGTCTTC	7420
	CCGCCCCAGG	TCCTACGCTG	AGTGCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCACGCT	GCTCTGCAGC	7490
	CTGTGCTACG	GCGACATGGA	GAACAAGCTG	TTTGGCGGGA	TTCCGCGGGA	CGGGTGAGGC	CTCTCTTCTC	7560
40	CCAGGGGGGG	TTGGGTGGGG	GTGATTTTGC	TTTTGATGCA	TTCACTGTTA	ATATTCCCTG	TGCTCTGGAG	7630
	ACCATGACTG	CTCTGTCTTG	AGGAACCCGA	CAAGGTTGCA	GGCCTTCTTT	GGTATGAAGC	CGCACGGGAG	7700
	GGGTGTCACA	GCCTGAGGAC	TGCGGGCTCC	ACCGAGGGCT	TGTCCAGCGG	CCATGTCCAG	AGGCCCTCAG	7770
	GCTCAGCAGG	CGGGAGGGGC	GCTGCCCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGGAGA	AGCACACAGG	7840
	CTTCTGTCTC	GTACCCACAG	TTCCGTTAGG	GTCTTGGGG	AGATGGGGCT	GGTGAGCGCT	GAGGCCCCAC	7910
45	ATCTGCCAGC	AGGCCCTCGA	CAGGTGGGCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCA	TCCCACTTGC	7980
	ATGGGGTCTA	CACCCAAAGG	CGCACACACC	TAAATATCGT	GCCAACTTAA	TGTGGTTCAA	CTCAGCTGCG	8050
	TTTTATTGAC	AGCAGTTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGGG	TACATGTGCA	CGACGTGAG	8120
	GTTAGTTTACA	TATGTATACA	TGTGCCATGT	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGTGT	8190
	TATCTCTCTA	TGCTATCCCT	CCCCACTCCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCCTCCAC	8260
	CTGTGTCCAA	GTGTTCTCAT	GTTCAGTTTC	CCACCTGTGA	GTGAGAACAT	GTGGTGTGTT	GTTCCTCTTC	8330
50	CTTGCAATAG	TTTGCTCAGA	GTGATGGTTT	CCAGCTTCGT	CCATGTCCCT	ACAAAGGACA	TGAACCTCAT	8400
	CTTTTTTATG	ACTGCATAGT	ATTCCGTGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTTGG	GTGGTTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCGCAAT	AAACATACGT	GTGCATGTGT	8540
	CTTTATAGCA	GCATGATTTA	TAATCCTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAAAATGGAT	8610
	TTCTAGTTCT	AGATCCTTGA	GGAATCACCA	CACGTGCTTC	CACAAATGGT	GAACTAGTTT	ACACTCCCAC	8680
55	CAACAGTGTA	AAAGTGTCTT	GGTGCTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTTATGA	AAATAGTATC	8750
	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGGT	TTTTCTGTG	CATCTTTTGA	AACCTTAGCT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTTCATAGAT	GAAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTGTGACA	8960
	CACAACTTGC	TCTGGGATTT	GGAGGAAAGT	GTCTCGAGC	TGGCGGCACA	CTGGTCAGCC	CTCTGGGACA	9030
	GGATACCTCT	GGCCCATGGT	CATGGGGCGC	TGGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGGCCAG	9100
60	CTTCTGTGG	ATAGGATCTG	GTCTCGGAT	CATGCTGAGG	ACCACAGCTG	CCATGCTGGT	AAAGGGCACC	9170
	ACGTGGCTCA	GAGGGGGCGA	GGTTCCAGC	CCCAGCTTTC	TTACCGTCTT	CAGTTATTTT	TTCCCTAAGAG	9240
	TCTGAGAAGT	GGGGCCGCGC	CTGATGGCCT	TGTTCTGTCT	TCAGCTGGCA	CAGAATTGCA	CAAGCTGATG	9310
	GTAACACATG	AGTACTTATA	ATGAATGAGG	AATTGCTGTA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTGCG	9380
	AAAGAAATTT	AAGTTTTCAT	TTTAAACCGCT	TTGGAGAATG	TTACTTTATT	TATGGCTGTG	TAAATGTTT	9450
65	GACATTACAT	CCCTCGTAGA	CAGATACTAC	GTAATAAGTG	TAAAGTTAAG	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCTGCGT	TTGGTGGATG	ATTTCTTGT	GGTGACACCT	CACCTCACCC	ACCGGAAAC	9590
	CTTCTCAGG	TGAGGCCCGT	GCCGTGTGTC	TGTGGGGACC	TCCACAGCCT	GTGGGCTTTG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTTGGC	ACCGCAGCGT	TGCTCTGCCC	AAGTCTCTCT	TCTCTGCCGG	TGCTGGATCC	9730
70	GCAAGAGCAG	AGGCCGTTGG	CCGTGCACCC	AGGCCCTGGG	GCGCAGGGGC	ACCTTCGGGA	GGGAGTGGGT	9800
	ACCGTGCAGG	CCCTGGTCTT	CGAGAGACGC	ACCCAGGTTA	CACACGTGGT	GAGTGCAGGC	GGTGACCTGG	9870
	CTCCTGTCTC	TCTTTGGAAA	GTCAAGAGTG	GCGGCTCCTG	GGGCCCCAGT	GAGACCCCCA	GGAGCTGTGC	9940
	ACAGGGGCTG	CAGGGCCGAG	GCGGCAGCCT	CCTCCCCAGG	GTGCACCTGA	GCCTGCGGAG	AGCAGAGCT	10010
	GCTGAGTGAG	CTGCCCCACA	CGGTTCCGCT	CGGTCACGTT	CCTGCGTGGG	GTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TTGCTGAGTG	CTGCTGTCTT	GAACCCAGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGTGAATC	AAACTAAAAA	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATTGTCCA	ATGTGGTCCC	10220
	CCTCAAGGGC	GGCCACACGA	GCCGGTGGGC	TGTTTTTAAA	GTGCGATTTC	ACGAGGGAGC	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCTT	CAGAAATAGT	GGCCGCCAGG	GGTGGTTTCA	GGTGCTTTGC	TGGGCTGTGT	10360
	TGTGAAAAAC	CCATTGGGAC	CGGCCCTCCA	AGTCCACCTT	CCAGTCCAC	CCTCCAGGGC	CGCCCTGGGC	10430



	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCCGAGC	ACAGCAGGCT	GTGCACATTT	AAATCCACTA	10500
	AGATTCACCT	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACGCTG	CTTCTGTGTG	GCAAGTTCTT	GAGGGTGCTG	GCCAGGGAGG	TGGCTCAGAG	10640
5	TGTATGTTGG	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTCGGC	AGCCATGTAA	CAGGAAGGGG	10710
	TGGCCACAGG	GAGCTGGGAA	TGCACCAGGG	GAGCTGCGCA	GCTGGCCGAG	GTCCCAGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCG	GGGCCACAGC	AGAGGCCGCA	GGAAAGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAGGCTACCG	GGCAGAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCCA	GGGAACCTCC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTTCGCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCCTG	AGCAGGAAGT	10990
10	CAGAACCCCT	CCCTTTGTCT	AAAGCACAGC	AGATGCCTTC	AGGGCATCTA	GGAGAAAACA	GGCAAAAGTCG	11060
	TTGAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTGTAGTCTG	CCCGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCCTGGCT	GTCCCGGGTC	CAGGCCAGGT	TCTTGATATG	TCACCTACCT	11270
	GTCTGCCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGGCTC	CTCAGAGCTC	11340
15	CTGCCAGGCC	CAGCACCTGT	CTCCAAATCA	CCACTTCTCT	GGGGTTTTCC	AAAGCATTTA	ACAAGGGTGT	11410
	CAGGTTACCT	CCTGGGTGAC	GGCCCCGCAT	CCTGGGGCTG	ACATTGCCCC	TCTGCCTTAG	GACCCCTGGT	11480
	CGAGGTGTCC	CTGAGTATGG	CTGCTGGTGT	AACTTGCGGA	AGACAGTGGT	GAACCTCCCT	GTAGAAGACG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCCGCCCA	CGGCCATTTC	CCCTGGTGGC	GCCTGTGTCT	11620
	GGATACCCCG	ACCCTGGAGG	TGCAGAGCGA	CTACTCCAGG	TGAGCGCACC	TGGCGGGAAG	TGGAGGCTGT	11690
20	GCCCGGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTTGCGTCC	ACCTCTGCTT	CCGTGTGGGG	CAGGCCACTG	11760
	CCAAATCCCA	AGGGTCAGAG	GCCACAGGGT	GCCCCCTGTC	CCATCTGGGG	CTGAGCAGAA	ATGCTCTTTT	11830
	CTGTGGGAGT	GAGGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAAAGGAA	ATGGTGCACC	11900
	AGACCTGGGT	GCACTGAGGT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AAGACGCCCG	GGCCCTGTCT	11970
	GGCGTGAGTC	TCTCAAACCC	GAACACAGGG	GCCCTGCTGG	GCATGAGTCC	CTCTGAACCC	GAGACCTCTG	12040
25	GGCCCTGTCT	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGCGGTGAG	TCTCTCCGCT	12110
	GTGAGCCCA	CACCTCAAGG	CTCATCCACA	GTCTACAGGA	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TCAGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTCCTC	CAGAGCCCCA	12250
	GAGCTCAAGG	CCCCGTCTCA	GGCTCAGACA	CAAAATGAAT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
	TGTTTCTTTT	ATGAATAAAA	AGTATCAACA	TTCCAGGCAG	GGCAAGGTGG	CTCACACCCTA	TAATCCCAGC	12390
30	ACTTTGGGAG	GCCGAGGTGG	GTGGATCACT	TGAGGCCCAGG	AGTTTGAGGC	CAACCTAACC	AACTAGTGA	12460
	AATTCATTTT	CTACTTAAAA	AATACAAAAA	TTAGCCTGGC	CTGGTGGCGT	ACGCCCTGAG	TCCCCGCTAT	12530
	CGGGGAGGCT	GAGGCAGGAG	AATCAATTGA	ACCCAGGAGG	CAGAGGTTCG	AGTGAGCCGA	GATCACACCA	12600
	CTGCACTCCA	GCCTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAATAAA	AAAAAGTATC	AGCATTTCTT	12670
	AACCATAGTG	GACAGGTGTT	TTTTTATTCT	GTCTTCGAT	AATATTTACT	GGTGTGTGTC	TAGAGGCCGG	12740
35	AACTGGGGGT	GCCCTCCTCT	GAAAGGCACA	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTGTGTTAA	12810
	CCAGAGGTTT	AAACTGGGGT	CCTGTGCTTC	TGAGTTAAAC	GTCCAGATCT	GGACTTTGCC	TCTTTCCAGA	12880
	ATGCTCCCTG	GGGTTTGCTT	CATGGGGGAG	CAGCAGGTGT	GGACACCCCT	GTGATGGGGG	AGCAGCAGGT	12950
	GCAGAGCGCC	TCATGATGGG	GGAGTGGGAG	GTGCAGACAC	CCTTGTGCAT	GGTGCCACAG	ATGTCCCTAT	13020
40	TGCAGTCCC	TCCCCACAAG	GATGCCGGTC	TGCTGTGCTC	CCCACAGTCC	CTGCTCCCTT	CTCACAGCTT	13090
	TACCTGGTCC	TGGCCTCCAC	TGGCTTTGTC	TGATGATATT	CCACATTTCC	TGGGCTCCCA	GCACCTCTTC	13160
	GCCTCTCCCA	GGCACCCTCT	CAGTGCTGGC	CATACCAGTC	AGCTGTGAAC	TGTCCACTGC	TTATTTTGCT	13230
	CCCCATGAAA	TGTATTTTTT	AGGACAGGCA	CCCTCGGTTT	CAGGCTCTGG	CACAGCATCA	GTGAATGTTA	13300
45	TTGAAGGACA	AAGGACAGAC	AAACAAATCA	GGAAAATGGG	TTCTCTCTAA	ACACATTGCA	AGGCCACAGA	13370
	GGGTAGTGCA	GGATGGGTGG	GCATCAGGTC	ATCAGATGTG	GGTCCAAATG	CAGAAATATC	TGTGCTCCCA	13440
	AAGGCCACTT	GGTCAGAGTG	TGTGCTTGCA	GAGGTGGCTC	TAAAAGTCTA	GCAGTGGAGG	CAGTGGTTCG	13510
50	CCATACTCAG	GGTGAATCA	CATCCTCTGT	GTCTGAAGTA	TACAGCAGAG	GCTTGAAGGG	CATCTGGGAG	13580
	AAGAAAACAG	GCAAAATGAT	TAAAGAAAGT	GAAAAAGGAA	AAGTGGTAAG	ATGGGAATTT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCAAAACCA	AGCTCAGATG	GTAGAATGTG	GTCAAGAACT	ATGGACAGAA	CAATAGAACCA	13720
	AAACGGAAGC	CCTATCTCTC	AGAAACGTGT	GTAAATGTGG	TATGTGGCAC	AGCTGATGGA	AAAGAGAGTG	13790
	TGTGTGTAAT	TTTTTTTCTT	GAGAAAACCTG	ACTGGAAGCA	AATAAGTTGT	GTCTTTACAG	CATATACCAG	13860
55	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAACAACA	CCAGCAACAG	AAATAAAACA	AAAGACTCAA	13930
	AGGGAAGGGA	GGTGAACGTT	CCCTGGTTTG	GTGTTGGGGA	AGGACACACA	GGGAGGGCGA	TGAACCAAGT	14000
	GAGGCAACGG	GCATTGCTTT	CACCTGCAGAG	AAACTCAGCT	TGCCTGAGCC	ACAGTGAAAA	TGGCCATTCC	14070
	CTGGAGCGTT	TGTGCACGTG	ATTTATTTAA	GGCGCCCTGT	GAGGTCTCTC	ACATTCATCC	TCTCACTTTG	14140
60	TTCTCTTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGGAGCAGC	CGCCCTTGGT	CACCCAGCTG	14210
	GCAAAAGGCA	TGCATGATTG	CAGCCTGGCC	TCTGTCTCCG	GGGCGCTTGC	TCTGCCCGAG	GACCCACAC	14280
	AAGTCAGACC	CATAGGCTCA	GGGTGAGCCG	GAGCCCAAGG	TCGTGTTGGG	GATGGCTGTG	AAAGAGGAAA	14350
	TGGAGCTCTG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAGAGAA	TGCATGTGAA	ACTGACAGCG	14420
	AGACCCATCC	CTCAAAGAAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
65	TTTTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTTGAGGCCAA	GCTGGAAAGA	CTTTCTGGA	AAGCAGCTTG	14560
	TTTGCATGGA	AGTCTCTACA	ATGCTCTGTG	TCTTCCAGT	AATTCACACT	CTGAAGTGAC	CAGACATTAT	14630
	CACGGGTCTT	ATTTACCAAT	TCCAGTGTTC	CAGGCAGGGG	GACTTGCCAC	AGCAAGTCAC	GAACTGCTCC	14700
	AAATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTT	CTCTGCCATT	AAACATTTTT	CAAAGATTTT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAACGTTT	ATTTCAATGT	AGCAGTGTTC	AAAGCTGGAT	GTAAGAAGAC	14840
70	ACACCCACAG	AGCCTGCCGT	GAATGTCATG	TGTGTTCACT	TTTGGACATG	GACATACATG	GGCAGTGAGT	14910
	GGTGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCTCCAT	CCTGCCCTCT	TGGAGACACC	ATGTTGGCCA	14980
	CGTGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTTCCATC	CCTGAGATTC	AAACACAGTG	15050
	AGATTCCCCA	CGCCCAACTC	AGTGTCTCC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCGAGGGAC	15120
	GCCCGGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTTTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
75	GATGATGAGT	GCACAAACAC	GGCCGTGGCA	GGTTTGGATA	CACCTAACAT	CACTAGCCAG	GTCTGGTGG	15260
	AGTTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCATTTGG	AGTCCATGGA	GTGAGCACCC	AGCCCGCTCG	15330
	GGCTGCAGCG	CATGCCCCAG	GCAGGACAA	GAAGCGGAGG	GAAGGCGAGG	GGCTCTTTGG	AGCAAGCTTT	15400
	GCAGGAGGGG	GCTGGGTGTG	GGGCAGGCAC	CTGTGTCTGA	CATTCCTCCC	TGTGTCTCAT	CTATGCCCGG	15470
	ACCTCCATCA	GAGCCAGTCT	CACCTTCAAC	CGCGGCTTCA	AGGCTGGGAG	GAACATGGCT	CGCAAACTCT	15540
	TTGGGGTCTT	GCGGCTGAAG	TGTCCAGGCC	TGTTTCTGGA	TTTGACAGGT	AGCAGGCTGA	TGTTAGGCAC	15610
	AGAGTTTCAGA	GTTCAGGAGG	TGTGTGCCCA	AGTATGTGTG	TGTGTGTGTG	CGCGCGTGCC	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACGTA	AGAGTGCACA	TGACGTCATA	TACACGTGAG	CACATACATG	TGTGCATGTG	15750
	TGTACATGAA	GGCATGGCAG	TGTGTGCACA	GGTGTGCAAG	GGCACAAGTG	TGTGCACATG	CGAATGCACA	15820



	CCTGACATGC	ATGTGTGTTT	GTGCACAGTC	GTGTGGGCGAT	TCACGTGAGG	TGCATGCGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGCTAG	GTCTCAGCA	15960
	CCAGTGCCAC	TCCTTACAGG	ATGAGACGGG	GTCCACAGCC	TTGGTGGGCT	GAGGCTCTGA	AGCTGCAGCC	16030
5	CTGAGGGCAT	TGTCCCATCT	GGGCATCCGC	GTCCACTCCC	TCTCTGTGG	GCTTCTGTGT	CCACTCCGCC	16100
	TCTCCTGTGG	GCATTACAT	CCACTCCACT	CCCTCTCTCC	TGTGGGCATC	CGCGTCCACT	CCCCCTCTCT	16170
	GTGGGCATCT	CGGTCCACCT	CCCTCTCTGT	TGGGCATTGT	CGTCCACTCC	CTCTCCTGGT	TCCTTCTCTG	16240
	CTTGGCCGAG	CCTCGGGGGC	AGGCAGATGA	CACAGAGTCT	TGACTCGCCC	AGGGTGGTTC	GCAGCTCGCC	16310
	GGTGAGGGCC	AGGCCGGAAT	TCACTGGGAA	GAGGGATAGT	TTCTTGTCAA	AATGTTCCCT	TTTCTTGTTC	16380
10	CATCTGAATG	GATGATAAAG	CAAAAAGTAA	AAACTTAAAA	TCCCAGAGAG	GTTTCTACCG	TTTCTCACTC	16450
	TTTCTTGGCG	ACTCTAGGTG	AACAGCCTCC	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCTGTGCA	16520
	GGCGTACAGG	TGAGCCGCCA	CCAAGGGGTG	CAGGCCACGC	CTCCAGGGAC	CCTCCGCGCT	CTGCTCACCT	16590
	CTGACCCGGG	GCTTCACTT	GGAACTCCTG	GGTTTATAGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTGC	16660
	TGCTGCTGTG	GCACAGTTCT	GTTCGGGTGG	CTCTGTGCAA	AGCACCTGTT	CTCCATCTCT	GGGTAGTGGT	16730
15	AGGAGCCGGT	GTGGCCCCAG	GTGTCCCCAC	TGTCCCTGTG	CAGTGGCCGT	GGGACGTCAT	GGAGGCCATC	16800
	CCAGGGCAGC	AGGGGCATGG	GGTAAAGAGA	TGTTTATGGG	GAGTCTTAGC	AGAGGAGGCT	GGGAAGGTGT	16870
	CTGAACAGTA	GATGGGAGAT	CAGATGCCCG	GAGGATTGGG	GGTCTCAGCA	AAGAGGGGCC	AGGTGGGTGC	16940
	AGGTGAGGGT	CGCTGGCCCC	ACCCCGGGGA	AGGTGACAGCA	GAGCTGTGGC	TCCCCACACA	GGCCGGCCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCCT	GGAACTTCC	CTGTCTGTGG	TGGTCAGGGG	GTGCCCTCTG	17080
20	CAAGAATCGA	CAACTTTATC	ACAGAGGGAA	GGGCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGGCTGG	17150
	AGTCAGGGCA	GGTGGTGCCA	CAAGCCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGCACCA	GGCCCGGGCC	17220
	TCCACCTCAA	CAGGCCCTCC	GAGCCACTGG	GAGCTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCCATGAGG	17290
	GCTGAGAAGG	AGTGTGAGCA	TTTGTGTTAC	CCAGGGCCGA	GGCTGCGCGA	ATTACCGTGC	ACACTTGATG	17360
	TGAATGAGG	TCGTGCTCTA	TCGTGGAAGG	CTCAGGGGAG	AGTTTTCAT	TACAAGGTGC	GTGCTGCTGC	17430
25	TACCATGAAA	ATGGTTTTTA	ACCCGAGTGC	TTCGCCCTTC	ATGCTCTGGC	AGGGAGGGCA	GAGCCACAGC	17500
	TGCATGTTAC	CGCCTTTGCA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCCAG	GGTGGCTCCG	17570
	GCTCAGACCG	CCCTCCTCTC	TGCCTTCTCT	CTCTGCCTCA	AATCTTCCCT	CGTTTGCATC	TCCTTGACGC	17640
	GTGCCCTGGC	CCTCGTGCAA	GCTGCTTAC	TCCTTTCGGG	AAACCCCTTG	GGTGTGCTGG	ATACAGGTGC	17710
	CAGTACAGG	TGGAGGTGTC	TGACACTGTG	GTTGACCCCA	GGGTCCAGCT	GGCGTGTCTG	GGGCTCCTTT	17780
30	GGGCCATGAT	GAGGTCAGAG	GAGTTTTCCC	AGGTGAAAC	TCCTGGGAAA	CTCCAGGGGC	CATGTGACCT	17850
	GCCACTGCT	CCTCCCATAT	TCAGCTCAGT	CTTGTCTCTA	TTTCCCCACC	AGGGTCTCTA	GCTCCGAGGA	17920
	GCTCCCGTAG	AGGGCCTGGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCCACC	CATGTGGGGA	CCCTTGGGTA	17990
	CTCGCTTGAT	TGGGTAGCCC	TGAGGAGGCC	GAGATGCGAT	GGGCCACGGG	CCGTTTCCAA	ACACAGAGTC	18060
35	AGGCACGTGG	AAGGCCCAGG	AATCCCTTTC	CCTCGAGGCCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCCGA	18130
	TTTCACGGCA	GCCAGGCTGC	AGTGGGCGAG	GCTGTGGTGG	TCCACGTGGC	GCTGGGGGCG	GGGTCTGATT	18200
	CAAAATCCGT	GGGGCTCGGC	CTTCTTGGCC	CGTGTGGGCC	CGGCCCTCAC	ACGGGCTTGG	TGTTGGACGC	18270
	CCGACCTCTA	GCAGGTGGCT	ATTCTCCCT	TTGGAAGAGA	GGCCCTCAC	CATGTAGGT	GTTTCCCTCC	18340
	TGGGTACGGA	CGGTGGCCGT	GTGGCAACCC	CGGGACCTTA	GGCTTATTTA	TTTGTTTAAA	AACATTCTGG	18410
40	GCCTGGCTTC	CGTTGTTGCT	AAATGGGGAA	AAGACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18480
	CGGGGTGCTG	GCTTGACTGG	TGTGATCTCA	GGTCATTCCA	GAAGTGGCTC	AGGAAGTCAG	TGAGACCCAG	18550
	TACATGGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGGCTCAG	GCAGTGGGTG	AGGCCAGGTA	18620
	CATGGGGGGC	TCAGGCATCG	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTCAG	ACAGAGGTACA	18690
	CGGGGGCTCT	GATCACACGC	ACATATGAGC	ACATGTGCAC	ATGTGCTGTT	TCATGGTAGC	CAGGTCTGTG	18760
45	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAAG	ATGGAGGCTG	ACAGGGCTGG	CGCGGTGGCT	18830
	CACACCTGTA	GTCCAGCAC	TTTGGGAGGC	CGAGGCCAGA	GGATCCCTTG	AGGCCAGGAG	TTTAAAGACA	18900
	GCCTGAGCAA	CATAGTAGAA	CCCCATCTCT	ATGAAAAATA	AAAACAAAAA	TTAGCTGAAC	ATGTTGCTGT	18970
	CGCCCTGTAG	TTCCAATACT	TGGGAGGCTG	AAGTGGGAGG	ATCACTTGAG	CCCAGGAGGT	GGAAAGCTGCA	19040
	GTGAGCTGAG	ATTGACCAC	TGTACTGCAG	CCTGGGTGAC	AGAGTGAGAG	CCCATCTCAA	CAACAACAAA	19110
50	GAAGACTGAC	AAATGCAGTT	TCTTGGAAAG	AAACATTAG	TAGGAACCTA	ACCTACACAC	AGAAGCCAA	19180
	TCGGTGTCTC	GGTGTCACTG	AGATGAGATG	ATGGGTCCCT	ACACCATCAC	CCCAGACCCA	GGGTTTATGC	19250
	ACCACAGGGG	CGGGTGGCTC	AGAAAGGATG	CGCAGGACGT	TGATATACGA	TGACATCAAG	GTGTCTGAGC	19320
	GAAGGGCAGG	ATTATGATA	AGTACCTGCT	GGTACACAAG	GAACAATGGA	TAAACTGGAA	ACCTTAGAGG	19390
	CCTTCCCGGA	ACAGGGGCTA	ATCAGAAGCC	AGCATGGGGG	GCTGGCATCC	AGGATGGAGC	TGCTTACAGC	19460
	TCCACATGCG	TGTTCATACA	GATGGTGAC	AGAAACGCAG	TGTACCTGTG	CACACACAGA	CACGAGCTA	19530
55	CTCGCACACA	CAAGCACACA	CACAGACATG	CATGCATGCA	TCCGTGTGTG	TGCACCTGTG	CCCATGAGGA	19600
	AACCCATGCA	TGTGCATTCA	TGCACGCACA	CAGGCACCCG	TGGGCCCATG	CCCACACCCA	CGAGCACCCG	19670
	CTGATTAGGA	GGCCTTCTCT	CTGACGCTGT	CCGCCATCCT	CTCAGGTTTC	ACGCATGTGT	GCTGCAGCTC	19740
	CCATTTCATC	AGCAAGTTTG	GAAGAACCCC	ACATTTTTC	TGCGCGTCAT	CTCTGACACG	GCCTCCCTCT	19810
	GCTACTCCAT	CCTGAAAGCC	AAGAAGCGAG	GTATGTGCAG	GTGCCGTGGC	TCAGTGGCAG	CAGTGGCTGC	19880
60	CTGCTGGTGT	TAGTGTGTCA	GGAGACTGAG	TGAATCTGGG	CTTAGGAAGT	TCTTACCCTT	TTTCCATGCA	19950
	GGAAGTGGTT	TAACCCAACC	ACTGTCAAGC	TCGTCTGCCC	GGCCTCTCGT	GGGGTGAGCA	GAGCACCTGA	20020
	TGGAAGGGAC	AGGAGCTGTG	TGGGAGCTGC	CATCCTTCCC	ACCTTGTCTT	GCCTGGGGAA	CGCCTGGGGG	20090
	GCCTGGTCTC	TCCTGTTTGC	CCCATGGTGG	GATTGTGGGG	GCCTGGCCCT	TCCTGTTTGC	CCTGTGGTGG	20160
	GATTGGGCTG	TCTCCCGTCC	ATGGCACTTA	GGGCCCTTGT	GCAAAACCCAG	GCCAAAGGGCT	TAGGAGGAGG	20230
65	CCAGGCCACG	GCTACCCAC	CCCTCTCAGG	AGCAGAGGCC	CGGTATCAC	ACGACAGAGC	CCCGGCCCGT	20300
	CCTCTGCTTC	CCAGTCACCG	TCCTCTGCCC	CTGGACACTT	TGTCCAGCAT	CAGGGAGGTT	TCTGATCCGT	20370
	CTGAAATTC	AGCCATGTGC	AACCTGCGGT	CCTGAGCTTA	ACAGCTTCTA	CTTTCTGTTT	TTTCTGTGTT	20440
	GTGGAAATTT	CACCTGGAGA	AGCCGAAGAA	AACATTTCTG	TCGTGACTCC	TGCGGTGCTT	GGGTCCGGAC	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCGT	CGGGTGTGGG	CAGCTTTCCT	GTGTCTCCTG	GGAGGGGAGC	20580
70	TGGGCTGGGC	CTGTGACTCC	TCAGCCTCTG	TTTTCCCCCA	GGGATGTGCG	TGGGGGCCAA	GGGCGCCGCC	20650
	GGCCCTCTGC	CCTCCGAGGC	CGTGCACTGG	CTGTGCCACC	AAGCATTCCT	GCTCAAGCTG	ACTCGACACC	20720
	GTGTCACTTA	CGTGCCACTC	CTGGGGTCA	TCAGGACAGG	CAAGTGTGGG	TGGAGGCCAG	TGCGGGCCCC	20790
	ACCTGCCACG	GGGTCACTCT	TGAACGCCCT	GTGTGGGGCG	AGCAGCCTCA	GATGCTGCTG	AAGTGCAGAC	20860
	GGCCCGGGGC	CTGACCTTGG	GGGCCTGGAG	CCACGCTGGC	AGCCCTATGT	GATTAAACGC	TGGTGTCCCC	20930
75	AGGCCACGGA	GCCTGGCAGG	GTCCCAACT	CTGCTTCCCA	TCTTGAACCC	TCTCAGGGGC	GATGGCTCCC	21000
	CACGCTTGGG	AGCCTTCTGA	CCCCGTGACT	GTGTCTCTCT	ACAGCCTCTT	CCCTGGCTGC	TGCCCTGAGC	21070
	TCTTGGGGTC	CTGAGCAAGT	TCTCTCCCCG	CCCCGCCGCT	CCAGCGTCAC	TGGGCTGCTG	CTGTGCTCGC	21140
	CCCGGTGGAG	GGGTGTCTGT	CCCTTCACTG	AGGTCCCCAC	CAGCCAGGGC	CACGAGGTGC	AGGCCCTGCC	21210

5 TGCCCGGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCTG 21280  
 ATTTTGGCCC CGCAGCCCAG ACGCAGCTGA GTCCGAAGCT CCCGGGGACG ACGCTGACTG CCCTGGAGGC 21350  
 CGCAGCCCAAC CCGGCACTGC CCTCAGACTT CAAGACCATC CTGGACTGAT GGCCACCCCG CCACAGCCAG 21420  
 GCGGAGAGCA GACACCAGCA GCCCTGTCCG CCGGGGCTCT ACGTCCCAGG GAGGGAGGGG CGGCCACAC 21490  
 CCAGGCCCGC ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT CCGGCTGAAG 21560  
 GCTGAGTGTC CGGCTGAGGC CTGAGCGAGT GTCCAGCCAA GGGCTGAGTG TCCAGCACAC CTGGCTCTT 21630  
 CACTTCCCCA CAGGCTGGCG CTCGGCTCCA CCCCAGGGCC AGCTTTTCTT CACCAGGAGC CGGGCTTCCA 21700  
 GCGCCACAT AGGAATAGTC CATCCCCAGA TTCGCCATTG TTCACCCCTC GCCCTGCCCT CTTTGGCTT 21770  
 10 CCACCCCCAC CATCCAGGTG GAGACCTTGA GAAGGACCTT GGGAGCTCTG GGAATTTGGA GTGACCAAAG 21840  
 GTGTGCCCTG TACACAGGCG AGGACCTTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGGAGGT 21910  
 GCTGTGGGAG TAAATACTG AATATATGAG TTTTTCAGTT TTGAAAAAAA TCTCATGTTT GAATCCTAAT 21980  
 GTGCACTGCA TAGACACCAC TGTATGCAAT TACAGAAGCC TGTGAGTGAA CGGGGTGGTG GTCAGTGGCG 22050  
 GCGCCATGCC TGGCTGTGCA TTTACGGGAG TCTATGAGTG AATGGGGTTG TGGTCAGTGC GGGCCCATGG 22120  
 CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190  
 15 GAGCCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGCCG GAAGGGCAGC AGGGATGCTG GGGGGCCAGC 22260  
 TTGGGGCGCG GGGATGATGG AGGGCTTGGC CAGGGTGGCA GGGATGATGG GGGCCCGCAG TGGGTGGGCA 22330  
 GGGGTGATGG GGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGAAAGCTG GCTGGGCCCC CTCTCCCTCT 22400  
 GCCTCCACCC TGCAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTTCTT 22470  
 GGAGGTGGGG GGCAGGGGCA TGACACCATC CTGTATAAAA TCCAGGATTC CTCTCTCTGA ACGCCCCAAC 22540  
 20 TCAGGTTGAA AGTCACATTC CGCCTCTGGC CATCTCTTAA AGAGTAGACC AGGATTCTGA TCTCTGAAGG 22610  
 GTGGGTAGGG TGGGGCAGTG GAGGGTGTGG ACACAGGAGG CTTAGGGGTG GGGCTGGTGA TGCTCTCTCA 22680  
 TCTCTTATC ATCTCCAGT CTCATCTCTC ATCTCTTAT CATCTCCAG TCTCATCTGT CTCTCTCTTA 22750  
 TCTCCAGTC TCATCTGTCA TCCTCTTACC ATCTCCAGT CTCATCTCTT ATCTCTTAT CTCTAGTCT 22820  
 25 CATCCAGACT TACCTCCAG GGGGGGTGCC AGGCTCGCAG TGGAGCTGGA CATACGTCCT TCTCTAGGCA 22890  
 GAAGGAACTG GAAGGATTGC AGAGAACAGG AGGGGCGGCT CAGAGGGACG CAGTCTTTGG GTGAAGAAAC 22960  
 AGCCCTCTCT CAGAAGTTGG CTTGGGCCAC ACGAAACCGA GGGCCCTGCG TGAGTGGCTC CAGAGCCTTC 23030  
 CAGCAGGTCC CTGGTGGGCG CTTATGGTAT GGGCGGGTCC TACTGAGTGC ACCTTGGACA GGGCTTCTGG 23100  
 TTTGAGTGCA CCGCGGACGT GCCTGGTGTC GGGGTGGGGG CTTATGGCCA CTGGATATGG GGTCAATTAT 23170  
 TGCTGCTGCT TCAGAGAATG TCTGAGTGAC CAGGCCTAAT GTGTATGGTG GGGCCAAAGT CACAGACTGT 23240  
 30 GTCGTAAATG CACTCTGGTG CCTGGAGCCC CCGTATAGGA GCTGTAGGGA AGGAGGGGCT CTTGGCAGCC 23310  
 GGCCTGGGGG CGCCTTTGCC CTGCAAACTG GAAGGGAGCG GCGCCGGGCG CCGTGGGCGG ACGACCTCAA 23380  
 GTGAGAGGTT GGACAGAACA GGGCGGGGAC TTCCAGGAG CAGAGGCCCG TGCTCAGGCA CACCTGGGTT 23450  
 TGAATCACAG ACCAACAGGT CAGGCCATTG TTCAGCTATC CATCTTCTAC AAAGCTCCAG ATTCTGTTT 23520  
 35 CTCCGGGTGT TTTTGTGTA AATTTTACTC AGGATTACTT ATATTTTTTG CTAAAGTATT AGACCCTTAA 23590  
 AAAAGGTATT TGCTTTGATA TGGCTTAACT CACTAAGCAC CTACTTTATT TGTCTGTTTT TATTTATTAT 23660  
 TATTATTATT ATTAGAGATG GTGTCTACTC GTCTACCCAG GTTGTAGTG CAGTGGCACA GTCATGGCTC 23730  
 GCGTGTAGCG CAACCCCA GGCCTCAAGT ATCTCCGGC CTCAGCTTCC CAGAGTGCTG GGATTACAGG 23800  
 40 TGTGAGCCAC TGCCCTTGCC TGGCACTTTT AAAAACCACT ATGTAAGGTC AGGTCCAGTG GCTTCCACAC 23870  
 CTGTATCCC AGTAGTTTG GAAAGCGAGG CAGAAGGATT GTCTAGGGCC AGGAGTTTGA GACCAGCATG 23940  
 GGTAACATAG GGAGACCCCA TCTCTACAAA AAATGCAAAA AGTTATCCGG GCGTGGGGTC CAGCATCTGT 24010  
 AGTCCAGCT GCTCGGAGG CTGAGTGGGA GGATCGCTTG AGCCCGGGAG GTCATGGCTG CAGTGAGCTG 24080  
 TGATTGTACC ATCGCACTCC AGCCTGGGCA ACAGAGTGAG ACCCTGTCTC AAAAAAAGG 24150  
 45 AAGGAGAGG AGAAGAGAAG AAGAGGAAG AAGGAAAGAG AAGAAGAAGG AAGAAGGAAG AAGAAGGAG 24220  
 AAGGAGGCT GCTAGGTGCT AGGTAGACTG TCAATCTCA GAGCAAAATG AAAATAACAA AGTTTTAAAG 24290  
 GGAAAGAAAA ACCCCAGCTC TTTGGACTTC CTTAGGCCGT AACTTCATCT CAAGCAGCTT CTTCCACAG 24360  
 ACAAGCGTG ATGGAGCGAG TGAGTTCAAA CGAGAAAGGG AGGAGAAGCA GGCAGGGGTG GAGGCTGTGG 24430  
 GTGACACCAG CCGAGACCCC TGAAGGGGAG TGGTTGTTTT CTTGCCCTCAG CCCCACGCTC CTGCCGCTC 24500  
 50 TGCACCTGCT GTAACGCTG ATGTTGGTGC CAGGTGCCCA CTTGGGAAGG ATGCTGTGCA GGGGGCTTGC 24570  
 CAAACTTTGG TGGGTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAATT ACAGCCCTC CCCCAGATG 24640  
 CCCAGCTCT TCTCTGGAA CCTGTGAATG TGTACCCCGC AAGGCAGAGG CTGGTGAAGG GTGAGGTGG 24710  
 AATCAGCGCT GCCAGTCAG CGATCTTAAG GTCATCTGG ATTATCTGTT GGGCCTGATA TGGCCACAAG 24780  
 GGTCCCTAGA AGTGAGAGAG GGAGGCGAGG GAGAGTCAGA GAGGGGACGT GAGAAGGACC ACTGGCCACT 24850  
 GCTGGCTTTG AGATGGAGGA GGGGGTCCCC AGCCAAGGAA TGGGGGCAGC CGCTCCATGC TGGAAAAGCA 24920  
 55 AGCAATCCTC CCCGGTCTG AGGGCACAGC GCGCTGCCCA CGGCTCGATT TCAGGCCAGT GGGACCTGTT 24990  
 TCAGCTTTCC GGCCTCCAGA GCTGTAAGAT GATGCGTTTG TGTTCAGCCA CTAAGCTGCA GTGATTCGTC 25060  
 ACAGCAGCAA ATGGAATAGC AGTACAGGGA AATGAATACA GGGACAGTTC TCAGAGTGAC TCTCAGCCCA 25130  
 CCGCTGGG

## 60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz  
 der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der  
 Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist  
 65 in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon				Intron				3'-Exon							
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk.(%)	70	60	80		100	100	95	70			80	100	100		60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position  
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

## Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCGGGGTCGGCGTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACCAGCGACATGCCGAGAGCAGCGCAGG  
CGACTCAGGGCGCTTCCCCCGCAG

5

## Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC  
CTGGTCCTCCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTCGAGTGGACACGGTGATCTCTGCC  
TCTGCTCTCCCTCCTGTCCAGTTTGCATAAACTTACGAGGTTACCTTACGTTTGTATGGACACGGGTTTCCAGGCGC  
10 CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGAGTGGAGCCGGTTGCGGCAATGGGAGAAGTGTCTGGAAG  
CACAGACGCTCTGGCGAGGGTGCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGGTGGGAATGAGAGGTGGGA  
CGAGAACCCCTCTTCTGGGGTGGGAGGTAAGGGTTTTCAGGTGCACGTGGTCAGCCAATATGCAGGTTTGTGTTTA  
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGTAATCCAGCACTTTGGGAAGCTGAGGCAGGTGGA  
TCACCTGAGGTGAGGAGTTGAGACCGCTGACCAACATGGTGAACCCCTATCTGTACTAAAAATACAAAAATTAGCTG  
15 GGCATGGTGGTGTGTGCCTGTAATCCAGTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC  
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAGAGTGAACTCTGTCTTTAAAAAAAAGTGT  
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCCTCTTTAGGTAT  
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGGTGGTTCAGGGG  
ATGGTGTCTGCTGGGCCCTGCCGTGTCCCCACCCTGTTTTCTGGATTGTATGTTGAGGAACCTCCGCTCCAGCCCCCTTT  
20 TGGCTCCCAGTGCTCCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTCACCCCTCCCCACAACTCCCAAGAC  
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTTCTTTTTTTCTTTTATGTTGGC  
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGCAGAATTGCTAACTCGGCGGTGTTTA  
CAGCAGGTTGCTTGAATGCTGCGTCTGCGTGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGCTGC  
GGCTCAGGTGGACCACGCCGAGTCAGATAAGCGTCATGAACCCAGTTTTGCTTTTTGTGCTCCAGCTTCTCTCGTTGAG  
25 GAGAGTTTGAGTTCTCTGATCAGGACTCTGCCTGTCTGCTGTTCTCTGACTTCAGATGAGGTGACAATCTGCCCTGG  
CTTATGCAGGGAGTGAGGCGTGGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT  
GTACAGTGTAGGGTGAGTGAGGCGCGGCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGGCCCCCGGGTGT  
CCCTGTACGTGTAGGGTGAGTGAGGCGCCATCCCGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGTGGTCCCGG  
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC  
30 CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCAGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCACC  
GTCCCTGGGTGTCCCTCCCAGGTATAGGGTGAGTGAGGCACTGTCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGG  
CGCGGCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT  
GAGGCTCTGTCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGTCTTTCTATAGCCACAGCT  
GCGCCGGTTGCCCATTTGCTGGGTAGATGGTGAGGCGCAGTGCTGGTCCCAAGCCTATCTTTCTGATGCTCGGCTCT  
35 TCTTGGTCACCTCTCCGTTCCATTTGCTACGGGGACACGGGACTGCAGGCTCTCGCTCCCGGTGCCAGGCACTGCAG  
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCTGGCTTGCTCACCACGTGCCCGCCACATGCATGCTGCCAATACTCC  
TCTCCAGCTTGTCTCATGCCGAGGCTGGAAGTCTGGGCTGCCTGTGTCTGCTGCCAGTGTGCTGGAGACATCCAGAA  
AGGGTTCTCTGTGCCCTGAAGGAAAGCAAGTCACCCCAGCCCCCTCACTTGCTCTGTTTTCTCCCAAGCTGCCCTCTGC  
TTGGCCCCCTTGGGTGGGTGGCAACGCTGTGCACCTATTCTGGGCACCTGCCGCTCATTTAGGCTGGGCTCTGCCT  
40 CCAGTCGCCCCCTCACATGGATTGACGTCAGGCCACAGGTTGAGTGTCTCTGTCTGTCTCTGCTCTGAGACCCAGCTG

GAGGGCCGGTGTCTCCGCCAGCCTTCGTGAGACTTCCCTCTTGGGTCTTAGTTTGAATTTCACTGATTACCTCTGACG  
TTTCTATCTCTCCATTGTATGCTTTTCTTGGTTTATTCTTTTCATTCTTTTCTAGCTTCTTAGTTTAGTCATGCCTTTC  
CCTCTAAGTGCTGCCTTACCTGCACCCTGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTCAAGTGTTCTTAAA  
ATACTTCAAAGTGTTAATACTTCTTTAAGTATTCTTATTCTGTGATTTTTTCTTTGTGACGCTGTGTTTTGACGTGA  
5 AATCATTTTGATATCAGTGACTTTTAAGTATTCTTTAGCTTATTCTGTGATTTCTTTGAGCAGTGAGTTATTGAACACT  
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT  
TGTGTAGTGGTCTGTATAATACCAATTATTGAAGTTGCGGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTCCAG  
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA  
AAGCTTCTGTCTCCTTCTAGATGCATGAAATCCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTTATT  
10 TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTATCTTCTGATGAGTGAA  
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATTCTTTTTTAAATTGCTCTTAGTACTGCCACACTGGGCTTCT  
TTTGATTAGTATTTTCTGCTGTGTCTGTTTTCTGCCCTTAAATTTATATATATATATATATTTTTTTTTTTTGGAGACA  
GAGTCTTGGTCTGTGCGCCAGGGTGAGTGAGTGGTGTGATCACAGGTGAGTGAACCTTTTACCTTCTGGCCTGAGCCGT  
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAACGACAGACACGACCGCTACACCTGGCTAATTTTAAATTTTTCTGGA  
15 GACAGGGTCTGTGTGTGTCAGGCTGGTCTCAAACTCTTGAGCTCAAGGATCCATCTACCTCGGCTTCCCAAAGTG  
CTGAATTACAGGCATGAGCCACCATGTCTGGCCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGTCTGTTA  
ACAGCATGTAGGTGAATTTCCAAATCCAGTCTGACAGTCGTTGTTTAACTGGATAACCTGATTTATTTTCAATTTTTTGTG  
ACTAGAGACCCGCTGGTGCATCTGATTCTCCACTTGCCTGTTGCATGTCCTCGTTCCCTGTTTCTCACCACCTCTTG  
GGTTGCCATGTGCGTTTCTGCGAGTGTGTGATCCTCTCGTTGCTCCTGGTCACTGGGCATTGCTTTTATTCT  
20 CTTTGCTTAGTGTTACCCCTGATCTTTTATTGTGCTGTTGCTTTGTTTATTGAGACAGTCTCACTCTGTCAACCA  
GGCTGGAGTGAATGGCACAATCTCGGCTCACTGCAACCTCTGCTCCTCGGTTCAAGCAGTTCTCATTCTCAACCTCA  
TGAGTAGCTGGGATTACAGGCGCCACCACACGCTGGCTAATTTTGTATTTTAGTAGAGATAGGCTTTCACCATGT  
TGCCAGGCTGGTCTCAAACTCCTGACCTCAAGTATCTGCGCGCTTGGCCTCCACAGTGCTGGGATTACAGGTGCAA  
GCCACCGTGCCCGCATACCTTGATCTTTTAAATGAAGTCTGAAACATTGCTACCCTTGCTCTGAGCAATAAGACCTT  
25 AGTGATTTTAGCTCTGGCCACCCCGAGCTGTGTGCTGTTTTCCCTGCTGACTTAGTCTATCTCAGGCATCTTGACA  
CCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTTGAGTGTCTGTAGCTTTGCCCCCGCTGCTTTCTCTC  
TTTGTTCCCCGTCTGTCTTCTGTCTCAGGCCCGCTGCGGTCCCTTCTTGTCTTTGCGTGGTCTTCTGTCTTG  
TTATTGCTGGTAAACCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGAGTCCGGGACCTCTGCTTATGATGC  
ACAGATGAAGATGTGGAGACTCACGAGGAGGGCGGTCTTGGCCCGTGAGTGTCTGGAGCACCAGTGGCCAGCGTTC  
30 CTTAGCCAGTGAGTGACAGCAACGTCCGCTCGGCCTGGGTTAGCCTGGAAACCCAGGCATGTGCGGGTCTGGTGGCT  
CCGCGGTGTCGAGTTTGAATCGCGCAACCTGCGGTGTGGCGCCAGCTCTGACGGTGTGCTGGCGGGGAGTGTCTG  
CTTCTCCTTCTGCTTGGGAACAGGACAAAGGATGAGGCTCCGAGCCGTTGTGCGCCACAGGAGCATGACGTGAGCC  
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCACGCCGTGAATCCAGCACTTTGGGAGGCCAAGCGGG  
TGGATCACGAGGTGAGGAGGTGAGACCATCTGGCCAACATGATGAAACCCCATCTGTACTAAAAACAAAAATTAGC  
35 TGGCGTGGTGGCGGGTGCTGTAATCCAGCTACTCGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGTTGGAA  
GTTGCAGTGAGCCGACATTGCACCACTGCCTCCAGCCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA  
AAAAAAAAAAATTCTAGTAGCCACATTAAAAAGTAAAAAGAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA  
GCCAGCATGTCCACACCTCATTTTTAGGGTGTATTGGTGGGAGCATCACTCACAGGACATTGACATTTTTTGAGC  
TTTGTCTGCGGGATCCCGTGTGTAGTCCCGTGTGCGCATCTCGGCCTGGACCTGTGGGCTTCCCATGGCCATGGCT  
40 GTTGTACCAGATGGTGCAGGTCCGGGATGAGGTGCGCAGGCCCTCAGTGAGCTGGATGTGCAGTGTCCGGATGGTGCAG  
TCTGGGATGAGGTGCGCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTGAGGGTGAAGTCTCCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGCAGGCCCTGCTGTGAGCTGGATG  
TGTGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTCTCCAGGCCCTCGGTAAGCTGGAGGTATGGAGTCCGGATGATGCA  
GGTCCGGGGTGAGGTCCGCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGGGTGAGGTCAAC  
AGGCCCTGCGGTGAGCTGGGTGTGCGGTGTCTGGATGGTGCAGGTCTGGAGTGGAGTCCGCAGACGGTGCCAGACCATGC  
5 GGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTGCCAGGCCCTGCTGTGAGTTGGATGTGGGGT  
GTCCGGATGTGTGCAGGTCCGGTGTGAGGTCAACAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCT  
GGGGTGAAGGTCCGCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGCAGGTCTGGAGTGGAGTCCGCAG  
GCCCTCGGTGAGCTGGATGTGCAGTGTCCAGATGGTGCAGGTCCGGGGTGAGGTCCGCAGACCCCTGCGGTGAGCTGGATG  
TGCGGTGTCTGGATGGTGCAGGTCTGGAGTGGAGTCCGCAGGCCCTCGGTGAGCTGGATGTATGGAGTCCGGATGGTGCC  
10 GGTCCGGGGTGAGGTCCGCAGACCCCTGCTGTGAGCTGGATGTGCGGTGTCTGGATGGTACAGGTCTGGAGTGGAGTCCGC  
AGACCCCTGCTGTGAGCTGGATATGCGGTGTCCGGATGGTGCAGGTCCAGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGGA  
GGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCGCAGGCCCTGCTGTGAACTGGATGTGCGGCGTCTGGATGGT  
GCAGGTCTGGGGTGTGGTCCGCAGGCCCTCGGTGAGCTGGAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTCCG  
CCAGGCCCTGCTGTGAGCTGGATGTGCGGCGTCTGGATGGTGCAGGTCTGGGGTGTGGTCCGCAGGCCCTCGGTGAGCTG  
15 GAGGTATGGAGTCCGGATGATGCAGGTCCGGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATG  
GTGCAGTCCGGGGTGAGGTCCGCAGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCTGGGGTGAGGT  
CACCAGGCCCTGCGGTGAGCTGGTTGTGCGGTGTCCGGTTGTGTCAGGTCCGGGGTGAGTTCCGCAGGCCCTCGGTGAGC  
TGGATGTGCGGTGTCCCCGTGTCCGGATGGTGCAGGTCCAGGGTGAGGTCCGCTAGGCCCTTGGTGGGCTGGATGTCCGT  
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGCAGGCCCTTGGTGAGCTGGATGTGCGGTGTCTGCATGGTGCAGGTCTG  
20 GGGTGAGGTCCGCAGGCCCTTGGTGGGCTGGATGTGTGGTGTCCGGATGGTGCAGGTCCGGCGTGAGGTCCGCAGGCCCT  
GCTGTGAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTAGCCAAGGCCCTTGGTGAGCTGGATGTGGG  
GTGTCCGGATGGTGCAGGTCCGGGGTGAGGTCCGCAGGCCCTGCGGTTAGCTGGATATGCGGTGTCCGGATGGTGCAGGT  
CCGGGGTGAGGTCAACAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCTGGATGGTGCAGGTCCGGGGTGAGGTCCGCAGG  
CCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGCAGGCCCTGCAGTGGAGCTGGATG  
25 TGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCCGCAGGCCCTGCGGTTAGCTGGATATGCGGTGTCCGGATGGTGC  
GGTCCGGGGTGAGGTCAACAGGCCCTGCGGTTAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGC  
AGGCCCTGCTGTGAGCTGGATGTGCTGTATCCGGATGGTGCAGGTCCGGGGTGAGGTCCGCAGGCCCTGCGGTGAGCTGG  
ATGTGCTGTATCCGGATGGTGCAGGTCTGGCGTGAGGTCCGCAGGCCCTGCGGTGAGCTGGATGTGCAGTGTACGGATGG  
TGCAGGTCCGGGGTGAGGTCCGCAGGCCCTGCGGTGGGCTGTATGTGTGTGTCTGGATGGTGCAGGTCCGGGGTGAGTT  
30 CGCCAGGCCCTGCGGTGAGCTGGATGTGTGGTGTCTGGATGTGCAGGTCCGGGGTGAGTTCCGCAGGCCCTCGGTGAGC  
TGGATATGCGGTGTCCCCGTGTCCGAATGGTGCAGGTCCAGGGTGAGGTCCGCAGGCCCTTGGTGGGCTGGATGTCCGT  
GTCCGGATGGTGCAGGTCTGGGGTGAGGTCCGCAGGCCCTTGGTGAGCTGGATGTGCGGTGTCCGGATGGTGCAGGTCCG  
GGGTGAGGTCAACAGGCCCTCGGTGATCTGGATGTGGCATGTCTTCTCGTTTAAG

35 Intron 3 (SEQ ID NO 6)

GTACTGTATCCCCACGCCAGGCCCTGCTTCTCGAAGTCTGGAAACACCAGCCCGGCCCTCAGCATGCGCCTGTCTCCACT  
TGCTGTGCTTCCCTGGCTGTGCAGCTCTGGGCTGGGAGCCAGGGGCCCGTCAAGGCCCTGGTCCAAGTGGATTCTGTG  
CAAGGCTCTGACTGCTGGAGCTCAGTTCTTACTTGTAAAATCAGGAGTTTGTGCCAAGTGGTCTCTAGGGTTTGT  
AAGCAGAAGGGATTTAAATTAGATGGAACACTACCACTAGCCTCCTTGCCTTTCCCTGGGATGTGGGTCTGATTCTCTC  
40 TCTCTTTTTTTTTTCTTTTTTGGATGGAGTCTCACTCTGTTGCCAGGCTGGAGTGCAGTGGCATAATCTTGGCTCACT

5 GCAACCTCCACCTCCTGGGTTTAAGCGATTACACAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC  
GCCTGGCTAATTTTTGTACTTTTAGGAGAGACGGGGTTTACCATTGTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG  
TGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTTAATT  
CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTAGGTCTAGAGAGGATAAAATCCACCCACTTGGCGACTCACTG  
10 CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTGAATG  
GCTGTGAGATTTGTCTGCAATGTTCTGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTGTCATCAGTGAG  
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCCGCCAGGC  
TGACTGTGGAGGGCTTAGTCAGAAGATCAGGGCTTCCCAGCTCCCTGCACACTCGAGTCCCTGGGGGGCCTTGTGAC  
ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG  
15 GACCCCGACGTGGTGTGGGGCCATTTCTTGCATCTGGGGGAGGGTCAGGGCTTTCCCTGTGGGAACAAGTTAATACAC  
AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAACATGGTCATTTGACCACTATTTTGAAAGAAT  
TTAATTGGGGTGACCGAAGGAGCAGACAGACGTGGTGGTCCCAAGATGCTCCTTGTCACTACTGGGACTGTTGTTCTG  
CCTGGGGGGCCTTGGAGGCCCTCTCCCTGGACAGGGTACCGTGCCCTTTTCTACTCTGTGGCCTGCGGCCTGCGGTC  
AGGGCACCAGCTCCGGAGCACCCGCGGCCAGTGTCCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCAGGT  
20 GTGGCCGCTCCAGCCCCGTGCCCCATGGGTGGTTTTGGGGGAAAAGGCCAAGGGCAGAGGTGTGAGGAGTGGTGGG  
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCCTCTCCATCTGAAGGGATGTGGCT  
CTTTCTACCTGGGGTCTGCTGGGGCCAGCCTTGGGTACCCAGTGGCTGTACCAGAGGGACAGGCATCCTGTGTGG  
AGGGGCATGGGTTCAGTGGCCCCAGATGCAGCCTGGGACCAGGCTCCCTGGTGTGATGGTGGGACAGTACCCTGGGG  
GTTGACCGCCGACTGGGCGTCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG  
25 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC  
TGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAGCTTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT  
CTATTGCAG

#### Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTTGGTTTAACTTCCTTTTTAAACAGAAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG  
CCCCGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCACGGCGCCAACCCATTGTGCGCACAGTGAAGTGGCCGAGG  
TGCCGGTGCTCCAGAAAAGCAGCGTGGGGGTGTAGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG  
CAGCCGGGCAGGGCCTGGATGCAGCACGGCCCGAGGTCTGGATCCGTGTCTGTGTGGTGCAGCCTCCGTGCGCT  
TCCGCTTACGGGGCCCGGGACAGGCCACGACTGCCAGGAGCCCACCGGGCTCTGAGGATCCTGGACCTTGCCCCACGG  
30 CTCCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCCGTCATCTGAGGAGAGTGTGGGGTGAAGTGGACAGAG  
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG  
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCTGGGTCCCTATGGTGGGGTGGGCAC  
TTGGCCGGATCCACTTTCTGACTGTCTCCCATGCTGTCCCGCCAG

#### 35 Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGGACCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG  
GAGGTACTCTGGGTGGGCGCAGGGAGTGAGGTGACCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC  
CTTCAGCCTTTCCTGCAGCACATGGGGCCGACTGTGCACCTGACTGCCGGGCTCCTATTCCCAAGGAGGGTCCCACTG  
GATTCCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCCGGTGCTTGACCCCACTCCTGAGCCAG  
40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGGACACAGCCCGCCCTGCCCTTGGGGTCTGGAGTGGTGGGGTTCAGAGAGAG





25

Intron 7 (SEQ ID NO 11)

30

35

40

## Intron 8 (SEQ ID NO 12)

GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGCGGGTGGCTGGGCGGGCTGGCAGGGCTTCTGCTCACCTCTCTCTGCCC  
CTTCCCCACTGNCCTTCTGCCCCGGGGCCACCAGAGTCTCTTTTCTGGCCCCCGCCCCCTCCGGCTCCTGGGCTGCAGGC  
5 TCCCGAGGCCCCGAAACATGGCTCGGCTTGGCGCAGCCGAGCGGAGCAGGTGCCACAGAGCCTGGAAATGGCAAGC  
GGGGTGTGGAGTTGCTCCTGCGTGGAGGACGAGGGGCGGGGGTGTGTCTGGGTGAGGTGTGCGCCGAGCGTTTGAGCCT  
GCAGCTTGTGAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCTCCGATACAAAA  
GGATTTTATCCGATTCTCATTCTGCTCCCTGTGCTGTGACCCCCGCGAGGGCGCGGGCTCTTCTCTCTGTGACTAGATTT  
CCCATCTGGAAGTGCGGGGTGACCGTGTAGTTTGTCTCTCTCGGGGGCCTGTGGTGGCCATGGGGCAGGCGGCTGG  
10 GAGAGCTGCCGTACACAGCCACTGGGTGAGCCACACTCACGGTGGTAGAGCCACAGTGCCCTGGTGCCACATCACGTCCT  
CTGGATTTTAAAGTAAACACACACCTCCCGGAGGCATCTGCCTGCGACCCCTGTGTGTGCTGGGGAGAGTGGTAGCAC  
GGAGGAAATTCGTGCACACTCAAGGTCATCAGCAAGGTCATCCGAGTCAGGTGGAACGTGGAGGCCTCTCTCTGGGATC  
GTCTCCAGCGGATAAAGGACTGTGCACAGCTTCGGAAGCTTTATTTAAAAATATAACTATTAATTATTGCAATTATAAGT  
AATCACTAATGGTATCAGCAATTATAATTTTATTAAGTATAATTAGAAATATTAAGTAGTACACAGTTCTGGAAAAA  
15 CACAAATTGCATGGCAGCAGAGTGAATTTGGCCGAGGACAGGTGTGCACATGTGTGAAGCGGCCCGAGGCCAC  
AGAATTCGCTGACAAAGTCACCTCCCGAGAGAACCCACCGGGCTCCTTCGTGGTGTGAATTTTATTAAGATGGATC  
AAGTCACGTACCGTCCAGTGTGGCAGGGCTTTGGGGAATGTGAGGTGATGACTGCGTCTCATGCCCTGACAGACAGGA  
GGTGACTGTGTCTGTCTGTCCCTAGGACACGGACAGGCCCGAAGCTCTAGTCCCCATCGTGGTCCAGTTTGGCCTCTGA  
ATAAAAACGTCTTCAAACCTGTTGCCCCAAAACTAAGAACAGAGAGAGTTTCCCATCCCATGTGCTCACAGGGGCGTA  
20 TCTGCTTGGTGTGACTCGCTGGGCTGGCCGGACTCCTAGAGTTGGTGTGCTGTCTCTGTGCAAAAAGTGCAGTCTCTCT  
GCCCCATCACTGTGATATCTGCACCAGCAAGGAAGCCTCTTTTCTTTCTTTTCTTTTCTTTTGGAGCGGAACGTCA  
CTGTGTCTGCTGGGCTTGTGAGTGCAGTGGCGCATCTCAACTCACTGCAACCTCCGCTCCCGGTTCCAGCATTCTC  
CTGCCCTCAGCCTCCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCCTGGCTAATTTTGTATTTTGTAGAGAG  
GGGTTTTTGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTG  
25 GGATTACAGGTGTGAGCCATCACGCCAGCCGGAAGCCTCTTTTAAAGTGACCACCTATAGCGCTTCCGAAAATAAC  
AGGTCTTGTTTTGCAGTAGGCTGCAAGCGTCTCTTAGCAACAGGAGTGGCGTCTGTGGGCTCTGGGATGGCTGAGGG  
TCGCGTGGCAGCCATGCCTTCTGTGTGCACCTTTAGGTTCCACGGGGCTATTCTGCTCTCACTGTTTGTCTGAAAACGCA  
CCCTTGGCATCCTTGTTTGGAGAGTTTCTGCTTCTCGTTGGTCACTGTGAAACTAGGGGCAAGTTGTATCCGTTGGCGC  
GCAGCGGCTACATGTAGGCTCATGAGTCTTTCACCGTGACAAATTCCTTGAACAAAAAAGGAGTCCGGTTAAGCAT  
30 TCATTCCGGGTCAAGTGTCTGGTTCTGTGAATAAACTCTAAGATTAAAGAAACCTTAATGAAAGAAAACCTTGATGATTC  
AGAGCAAGGATGTGGTACACCTGTGGCTGGATCTGTTTCAGCCGCCCCAGTGCATGGTGAAGTGGGAGCAGGGATTG  
TTTGTTCAGAGGTCTCATCTGGTATGTTTCTGAGGTGTTTCCCGCTGAATGGTAGACGTGTGCTTGTGTGTATGAGGT  
TCTGTGTCTGTGTGGCTCGGTTTGTAGTGTACGCATGTCCAGCACATGCCCTGCCGTCTCTCACCTGTGTCTTCCCGC  
CCAG

35

## Intron 9 (SEQ ID NO 13)

GTGAGGCTCCTCTTCCCGAGGGGGCTTGGGTGGGGGTGATTGCTTTTGATGCATTCACTGTTAATATTCTGGTGC  
TCTGGAGACCATGACTGCTCTGTCTTGAGGAACAGACAAGGTTGCAGCCCCCTCTTGGTATGAAGCCGACGGGAGGGG  
TTGCACAGCCTGAGGACTGCGGGCTCCACGCAGGCTCTGTCCAGCGCCATGTCCAGAGGCCTCAGGGCTCAGCAGCGG  
40 GAGGGCGCTGCCCTGCATGATGAGCATGTGAATTAACACCGAGGAAGCACACCAGCTTCTGTCACTCACCCAGGTTT

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGCAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA  
CTGGGCGCCTCTTCAGCCCATGCCCCTCCACTTGCATGGGGTCTACACCAAGGAAGCACACACCTAAATATCGTGCC  
AACCTAATGTGGTTCAACTCAGCTGGCTTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC  
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCATCATTTACA  
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCTG  
TGTCCAAGTGTTCTCATGTTCAGTTCACCTGTGAGTGAGAACATGTGGTGTGGTGTTCCTTTCCTTGCATAGTTT  
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCTTTTTTATGACTGCATAGTATT  
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTTGGTTGCAAGTCTTTGCTACT  
GTGAATAGTGCCGCAATAAACATACGTGTGCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGGGTATATACCCAG  
10 TAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCACAATGGTTGAA  
CTAGTTTACACTCCACCAACAGTGTAAAAGTGTCTGGTGTGGAGAGGATGTGGACAGCAGTTATTTTTTATGAAAA  
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTCTCTCGAAGAC  
TCCGGGTTTTTCTGTGCATCTTTGAAACTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA  
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAACAGAAACAAAATTTCTTGACACACAACCTGCTCTGGGATTGGA  
15 GGAAAGTGTCCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG  
GCTTGGGCTGAGGGTCACACAGTGCCATGCCCAGCTTCCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC  
ACAGCTGCCATGTGGTAAAGGGCACACGTGGCTCAGAGGGGCGAGGTTCCAGCCCCAGCTTCTTACCGTCTTCAG  
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCCTTCGTTCTTCAGCTGGCACAGAATTGCACAA  
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTGGAAA  
20 GAAATTTAAGTTTTTCATTTAACCCTTTGGAGAATGTTACTTTATTTATGGCTGTGTAAATTGTTTGACATTCACTCCC  
TCGTAGACAGATACTACGTAAAAAGTGTAAAGTTAACTTGCTGTGATTTTCCCTTATTTTAG

## Intron 10 (SEQ ID NO 14)

GTGAGGCCCCGTGCGGTGTCTGTGGGGACCTCCACAGCCTGTGGGCTTGCAGTTGAGCCCCCGTGTCTGCCCCCTGG  
25 CACCGCAGCGTTGTCTCTGCCAAGTCTCTCTCTGCGGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC  
CAGGCTCGGGGCGCAGGGGCACCTTCGGGAGGGAGTGGGTACCGTGCAGGCCCTGGTCTCTGAGAGACGCACCCAGGTT  
ACACACGTGGTGAAGTGACAGGCGGTGACCTGGCTCCTGCTGCTCTTTGGAAAGTCAAGAGTGGCGGCTCTGGGGCCCCAG  
TGAGACCCCCAGGAGCTGTGCACAGGGCCTGCAGGGCCGAGGCGGCAGCCTCCTCCCAGGCTGCACCTGAGCCTGCGGA  
GAGCAGGAGCTGTGAGTGAGCTGGCCACAGCGTTGCTGCGGTACGTTCTGCGTGGGGTTGTTGGGATCGGTGGG  
30 AGAATTTGGATTTGCTGAGTGCTGCTGCTTGAACACGGAGATGGCTAGGAGTGGGTTTCAGAGTTGATTTTGTGAAT  
CAAATAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG  
AGCCGGTGGGCTTGTTTTAAAGTGCATTTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGGAACCTCAGAAAAATG  
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTTGTGAAAACCATTTGAGACCGCCCTCCAAGTCCACCC  
TCCAGGTCCACCCCTCAGGGCCGCGCTGGGCTGGGGGTATGCTGGCGTTCCTGTGCGGCAGCCCGAGCACAGCAGGC  
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAGCAACTGAGGGCTCAGGAGTCTGAGGCT  
GCTGAGGGGACAGAGCAGACGGGAACGCTGCTTCTGTGTGGCAAGTTCTGAGGGTGTGGCCAGGGAGGTGGCTCAGA  
GTGTATGTTGGGGTCCACCGGGGGCAGAACTCTGTCTCTGATGAGTCGGCAGCCATGTAACAGGAAGGGGTGGCCACAG  
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCGAGGTCCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCCC  
GGGGCCACAGCAGAGGCCGAGGAAGGGAAGGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGCTCCCTG  
40 AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCCAGCTCACAG

CCCAGCCAGGTCCCGCGCTGAGCAGGAACTCAGAACCCCTCCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT  
AGGAGAAAACAGGCAAAGTCGTTGAGAAACGTCTTAAAGAAGGTGGGATGGTGGCAATTTCTGTCCAGATTTTAGTCT  
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC  
TGGGGCTGCACCTCCCATCTGAGTCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGATGCTCACCTACCTGTCTCTGCC  
5 GGGAGACAGGGAAGCACCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCACAGCCCT  
GCTCCAAATCACCACCTTCTCTGGGGTTTTCCAAAGCATTTAAACAAGGGTGTGAGGTACCTCCTGGGTGACGCCCCGCA  
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

# Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGAAGTGGAGCCTGTGCCCCGGCTGGGGCAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCT  
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAAGGGTCAGAGGCCACAGGGTCCCCCTCGTCCCATCTGGGGCTGAGCAGA  
AATGCATCTTTCTGTGGGAGTGAGGGTGTCTACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAAAAGGAAATGGTGAC  
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCAAGACGCCCCGGGCCCTGCTGGGCGTGAGT  
CTCTCAAACCCGAACACAGGGGCCCTGTGGGCATGAGTCCCTCTGAACCCGAGACCCTGGGGCCCTGCTGGGCGTGAGT  
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC  
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA  
AATTCTGGGGTCTTGTTCCTCCAGAGCCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA  
CACAGATGCAGAAATCTGTGCTGTTTCTTTATGAATAAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT  
ATAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG  
20 AAATTCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGGCACACGCTGTAGTCCCCGCTATGCGGGAGGC  
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTGTCAGTGAGCCGAGATCACACCACTGCACTCCAGCCTGGGCA  
ACAGAGTGAGACTTCATCTTAAAAAAGGATATCAGCATTCCAAAACCATAGTGACAGGTGTTTTTTTATTTC  
TGTCCTTCGATAATATTACTGGTGTGTGCTAGAGGCCGGAACCTGGGGGTGCCTTCTCTGAAAGGCACACCTTCATGG  
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACCAGAGGTTTAAACTGGGGTCTGTGCTCTGAGTTAAGTCCAGATC  
25 TGGACTTTGCTCTTTCCAGAAATGCTCCCTGGGGTTTGCTTCATGGGGGAGCAGCAGGTGTGGACACCCTCGTGATGGGG  
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGGAGTGGCAGGTGCAGACACCCTTGTCATGGTGCCAGCATGTCCTGT  
TTGCAGCTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTCCCTCTCACAGCCTTACCTGGTC  
CTGGCCTCCACTGGCTTTGTCTGCATGATTTCCACATTTCTGGGCTCCAGCACCTCTTCGCCTCTCCAGGCACCTCT  
GCAGTGCTGGCCATACCACTCAGCTGTGAAGTGTCCACTGCTTATTTTGTCTCCCATGAAATGTATTTTTTAGGACAGGC  
30 ACCCTGGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAATGG  
GTTCTCTTAAACACATTGCAAAGCCACAGAGGCTAGTGCAGGATGGGTGGGCATCAGGTATCAGATGTGGGTCCAATG  
CCAGAAATATTCTGTGCTCCCAAAGGCCACTTGGTCAGAGTGTGTGCTTGAGAGGTGGCTCTAAAAGCTCAGCAGTGGAG  
GCAGTGGTTGCCATACTCAGGGTGAATCACATCCTCTGTGCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA  
GAAGAAAACAGGCAAAATGATTAAGAAAAGTGAAGGAAAGTGGTAAGATGGGAATTTCTTGTCCAGATTTTAGTC  
35 TCCCAAACACAGCTCAGATGGTAGAATGTGGTCAGAACTGATGGACAGAACATAGAACAAAACGGAAGCCCTATCTCT  
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGTGTGTGTGAATTTTTTTTCTGAGAAAAC  
GACTGGAAGCAAATAAGTTGTGCTTTACAGCATATACAGAGCAGATTCTAGGTAGAAGAGGAGACACATGCAAAACAC  
ACCAGCAACAGAAATAAAACAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC  
AGGGAGGCGGATGAAACAGTGAGGCAACGGGCATTGCTTTCACTGCAGAGAACTCAGCTTGCTGAGCCACAGTGAAA  
40 ATGGCCATTCCCTGGAGCGTTTGTGCACGTGATTTATTTAAGCGCCCTGTGAGGTCTGCACATTATCCTCTCACTTT

5 GTTCTCCTAACCACTGAGAGGTAGAGGAGGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC  
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCCGAGGACCCACACAAGTCAGACCCATAGGCTC  
AGGGTGAGCCGGAGCCCAAGGTCGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC  
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC  
10 GAGACCTGTCCCCATCCCTCATGCTGGCTCCTTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAG  
ACTTTTCTGGAAGCAGCTTGTTTGCATGGAAGTCTCACAATGTCTGTCTTCCAGTAATTCACCTTCTGAAGTGA  
CCAGACATTATCACGGGTCTTATTTACCATTTCAGTGTTCAGGCAGGGGGACTTGCCACAGCAAGTCACGAACCTGCC  
CAAATACAGGGCTAAGGAGATATTATGCATCAGAACTTGCTCTGCCATTAAACATTTTCAAAGAATTTTGAAGAAT  
GTTTAATGGCAGAAACGTTTATTTCAATGTAGCAGTGTCAAAGCTGGATGTAAAGAACAACCCCCAGGAGCCTGCCG  
15 TGAATGTCTGTGTTCATCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGGAGGCCCTGGAGGACATCGGTGG  
GATGCCCTCCATCCTGCCCTCTGGAGACACCATGTGTGCCACGTGCACTCACTGGAGCCCTGTTTAGCTGGTGCCACCTG  
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATCCCCACGCCCACTCAGTGTCTCCCAAAAAACCTGAGTCAC  
ACCTGTGTTCACTCGAGGGACGCCCGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTGGCTGAGTTATGTGCAGATC  
TCATCAGGGCAGATGATGAGTGCACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCTGGTG  
20 GAGTTTGGTCTATGCAGAGTCTGGATGGCATGTAGCATTTGGAGTCCATGGAGTGAGCAGCCAGCCCTCGGGCTGCAGC  
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTCAGGAGGGGGCTGGGTGT  
GGGGCAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

#### Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCACAGAGTTCAGAGTTCAGGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCGT  
GCCTGCAAGGCTGATGGTGAAGTGCACGTAAAGAGTGACATGTACGCATATACAGTGAGCAGATACATGTGTGCAT  
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCATGCGAATGCACACCTGACA  
TGATGTGTGTTCGTGCACAGTGTGTGGGCATTACGTGAGGTGCATGCGTGTGGGTGTGAGTGTGAGTAGCATGTGT  
GCACATAACATGTATTGAGGGGTCTCGTGTTCACCCCGCTAGGTCTCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC  
25 GGGGTCCCAGGCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCCCTCTGGGCATCCGCGTCCACT  
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTTACATCCACTCCACTCCCTCTCTCTGTGGGC  
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACCTCCCCCTCTCTGTGGGCATTGCGTCCACTCCCTCTCCT  
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGGCAGGCAGATGACACAGAGTCTTGACTCGCCAGGGTGGTTCGCAGCTG  
CCGGGTGAGGGCCAGGCCGATTCTACTGGGAAGAGGGATAGTTTCTGTCAAATGTCTCTTCTTCTTCTTCCATCTGA  
30 ATGGATGATAAAGCAAAAGTAAAACTTAAATCCCAGAGAGGTTTCTACCGTTTCTCACTCTTCTTGGCGACTCTAG

#### Intron 13 (SEQ ID NO 17)

35 GTGAGCCGCCCAAGGGGTGCAGGCCAGCCTCCAGGGACCCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACCT  
TGGAACCTCTGGGTTTATAGGGCAAGGAATGTCTTACGTTTTTCAGTGGTGTCTGCTGTGCACAGTCTGTTCGCGTG  
GCTCTGTGCAAAGCACCTGTTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCCTGT  
GCACTGGCCGTGGGACGTATGGAGGCCATCCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG  
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGGAGGATTGGGGTCTCAGCAAAGAGGGCC  
GAGGTGGGTGCAGGTGAGGTCGCTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGTCCCCACACAGCCGGCCA  
GCACCTGTGCTCTGGGCATGGCTGTCTCTGGAACGTTCCCTGTCTGTGGTGGTCAAGGGGTGCCCTGCCAAGAATCG  
40 ACAACTTTATCACAGAGGGAAGGGCCAATCTGTGGAGGCCACAGGGCCAGCTTCTGCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGCTGTACCAAAGGCGAGTCGGGCACACAGGCCCGGGCTCCACCTCAACAGGCCTCCGAGCCACTG  
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTACCCAGGGCCG  
AGGCTGCGCGAATTACCGTGCACTTGATGTGAAATGAGGTCGTCTATCGTGGAACCCAGCAAGGGCTCACGGGA  
GAGTTTTCCATTACAAGGTCGTACCATGAAATGGTTTTTAACCCGAGTGCTTGCGCCTTCATGCTCTGGCAGGGAGGGC  
5 AGAGCCACAGCTGCATGTTACCGCCTTTGCACCAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCAGGGTGCCTCC  
GGCTCAGACCGCCTCCTCTCTGCTTCTCTCTCTGCTCAAATCTTCCCTCGTTTGCATCTCCCTGACGCGTGCTGGG  
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGAAACCTTGGGGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT  
CTGACACTGTGGTTGACCCAGGGTCCAGCTGGCGTGCTTGGGGCCTCCTTGGGCCATGATGAGGTGACAGGAGTTTTCC  
CAGGTGAAAACCTCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCCTCCCATATTAGCTCAGTCTTGTCTC  
10 ATTTCCCCACCAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCCAC  
CCATGTGGGGACCCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA  
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGGAGAGCTGGGCCCCG  
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGGCGGGGTCTGATTCAAATCCGC  
TGGGGCTCGGCCTTCTGGCCCGTGCTGGCCGCGCTCCACACGGGCTTGGGGTGGACGCCCCGACCTCTAGCAGGTGGC  
15 TATTTCTCCCTTGGGAAGAGAGCCCTCACCATGTAGGTGTTTCCCTCCTGGGTGAGGAGCGTGGCCGTGTGGCAACC  
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCTGGCTTCCGTTGTGCTAAATGGGAAAGACATCC  
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGGTGCTGGCTTACTGGTGTGATCTCAGGTCAATCCAGAAGTGGCT  
CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGTGGGT  
GAGGCCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGAGGTCAGACCAGGTAC  
20 ACGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC  
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCCAGCA  
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCCATCTC  
TATGAAAAATAAAAAAAAATTAGCTGAACATGGTGGTGTGCGCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG  
GATCACTTGAGCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGCAGCCTGGGTGACAGAGTGAGA  
25 GCCCATCTCAACAACAACAAGAAGACTGACAAATGCAGTTTCTGGAAAGAAAATTTAGTAGGAACCTTAACCTACACA  
CAGAAGCCAAAGTCGGTGTCTCGGTGTGAGTGAGATGAGATGAGGTGCTCACACCATCACCCAGACCCAGGGTTTATG  
CACCAAGGGGCGGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGGTTGTCTGACGAAGGGCAG  
GATTCATGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAACCTTAGAGGCCTTCCCGAACAGGGGCT  
AATCAGAAGCCAGCATGGGGGCTGGCATCCAGGATGGAGCTGCTTCAGCCTCCACATGCGTGTTCATACAGATGGTGCA  
30 CAGAAAACGAGTGTAACCTGTGCACACACAGACACGAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC  
ATCCGTGTGTGTGCACCTGTGCCATGAGGAAACCATGTCATGTGCATTATGCACGCACACAGGCACCGGTGGGCCAT  
GCCACACCCACGAGCACCGTCTGATTAGGAGGCTTTCTCTGACGCTGTCCGCCATCCTCTCAG

#### Intron 14 (SEQ ID NO 18)

GTATGTGCAGGTGCCTGGCCTCAGTGGCAGCAGTGCCTGCCTGCTGGTGTAGTGTGTCAGGAGACTGAGTGAATCTGGG  
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCCAACCACTGTGAGGCTCGTCTGCCCGCCCTCTCGT  
GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCACCTTGCTCTGCTGGGGAA  
GGCTGGGGGGCTGGTCTCTCTGTTTGGCCCATGGTGGGATTGGGGGGCTGGCCTCTCTGTTTGGCCCTGTGGTGG  
GATTGGGCTGTCTCCCGTCCATGGCACTTAGGGCCCTTGTGCAAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG  
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCTATCACACGACAGAGCCCGCGCGCTCTGCTTCCAGTCACCG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTGGAACCTGCGGT  
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTCTGTGTTGTGGAAATTTACCTTGAGAAGCCGAAGAAAACATTTCCTG  
TCGTGACTCCTGCGGTGCTTGGGTCCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTCCG  
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

5

**Intron 15 (WEQ ID NO 19)**

GCAAGTGTGGGTGGAGGCCAGTGCGGGCCCCACCTGCCAGGGGTCTCCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC  
AGATGCTGCTGAAGTGACAGCGCCCCCGGGCCTGACCTGGGGGCTTGAGCCACGCTGGCAGCCCTATGTGATTAAACG  
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAACTTCTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC  
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGCTCTCACAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGT  
CCTGAGCAAGTTCTCTCCCCCGCCGCTCCAGCGTCACTGGGCTGCCTGTCTGCTCGCCCCGGTGGAGGGGTGTCTG  
TCCCTTCACTGAGGTTCCACACAGCCAGGGCCACGAGGTGCAGGCCCTGCCTGCCCCGCCACCCACAGCTCTAGGAGGG  
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTGGCCCCGAG

10

**3'-untranskribierter Bereich (SEQ ID NO 20)**

ATCTCATGTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT  
GGTCAGTGCGGGCCCATGGCCTGGCTGTGCATTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCGGGCCCATG  
GCCTGGCTGGGCTGGGAGGTTTCTGATGCTGTGAGGCAGGAGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA  
CCCTGGAAGACATAACAGTAAGTCCAGGCCGAAGGGCAGCAGGGATGCTGGGGGCCAGCTTGGGCGCGGGGATGATG  
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGGCCCAGCTGGGGTGGCAGGGGTGATGGGGGGGGCTGGTCTGGGTGG  
CGGGGAAGATGGGAAGCCTGGCTGGGCCCCCTCTCCCTGCCTCCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT  
GGTGACATCCTCTGGGCCATCAGCTTTCTGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT  
CCTCCTCCTGAACGCCCAACTCAGGTTGAAAGTCACATTCCGCCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG  
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAAGGTGGGGCTGGTGATGCTCTCTC  
ATCCTCTTATCATCTCCAGTCTCATCTCTCATCCTTATCATCTCCAGTCTCATCTGTCTTCTCTTATCTCCAGT  
CTCATCTGTCTATCTTACCATCTCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA  
GGGCGGGTGGCAGGCTCGCAGTGGAGCTGGACATACGTCCTTCTCAGGCAGAAGGAAGTGAAGGATTGCAGAGAACAG  
GAGGGCGGCTCAGAGGGACGAGTCTTGGGTGAAGAAACAGCCCTCCTCAGAAGTTGGCTTGGGCCACAGAAACCG  
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGCTCTACTGAGTG  
CACCTTGGACAGGGCTTCTGGTTTGAGTGAGCCCCGACGTGCCTGGTGTGGGGTGGGGCTTATGGCCACTGGATATG  
GCGTCATTTATTGCTGTGCTTCAAGAAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCCAAGTCCACAGACTG  
TGTGTAATGCACTCTGGTGCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCCTGGGG  
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGCCCCGGGCGCGTGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC  
AGGGCGGGGACTTCCAGGAGCAGAGGCCGCTGCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT  
GTTCACTATCCATCTTCTACAAAGCTCCAGATTCTGTTTCTCCGGGTGTTTTTTGTTGAATTTTACTCAGGATTACT  
TATATTTTTTGCTAAAGTATTAGACCCTTAAAAAGGTATTGCTTTGATATGGCTTAACCTACTAAGCACCTACTTTAT  
TTGTCTGTTTTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTACCCAGGTGTTAGTGCAGTGGCAC  
AGTCATGGCTCGCTGTAGCCGCAACCCCGAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCAGAGTGTGGGATTACAG  
GTGTGAGCCACTGCCCTTGCCTGGCACTTTTAAAAACCACTATGTAAGGTCAGGTCCAGTGGCTTCCACACCTGTCTATCC  
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAACCATGAGGAGACCC

20

25

30

35

40



ATCTCTACAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG  
AGGATCGCTTGAGCCCGGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA  
GACCCTGTCTCAAAAAAAAAAAAAAAAAAAGAGGAGAAGGAGAAGAGAAGAAGGAAGAAGGAAAGAGAAGAAGAAG  
5 GAAGAAGGAAGAAAGAAGGAGAAGGAGGCTGTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAATAACA  
AAGTTTAAAGGGAAAGAAAAACCCAGCTCTTTGGACTTCCTTAGGCCTGAACTTCATCTCAAGCAGCTTCCTTCCACA  
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA  
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCCTCAGCCCCACGCTCCTGCCGGTCTGCACCTGCTGTAACCGTC  
GATGTTGGTGCCAGGTGCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG  
10 GCACTTGTGGCAGGCACAATTACAGCCCTCCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTCAACCG  
CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCCTGGATTATCTGG  
TGGGCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC  
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAGGC  
AAGCAATCCTCCCGGTCCTGAGGGCACACGGCCCTGCCACGCTCGATTTAGGCCAGTGGGACCTGTTTCAGCTTTC  
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTTCAAGCTGAGTGCAGTGATTTCGTCACAGCAGCAATGGAATAG  
15 CAGTACAGGGAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorenggeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

#### Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA<sup>+</sup>-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur  
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv  
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8×10<sup>5</sup>  
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl  
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),  
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl  
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser  
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die  
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach  
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden  
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl  
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-  
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min  
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol  
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stopplösung gelöst. Nach 5  
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel  
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom  
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in  
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA<sub>-1</sub>TTGT)  
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA<sub>n</sub>Na/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei  
10 anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

#### Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20

Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem  
25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für  
30 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprinten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

### Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Spl-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

### Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTGCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die  
5 verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und  
10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 $\alpha$ -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für  
15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

#### Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection  
25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

30



HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wasssergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO<sub>2</sub> kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit  
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils 2x 10<sup>5</sup> –HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium  
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium  
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H<sub>3</sub>PO<sub>4</sub>; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl  
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO<sub>3</sub>)<sub>4</sub>; Mg(OH)<sub>2</sub>·5H<sub>2</sub>O; 2,67 mM MgSO<sub>4</sub>; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl<sub>2</sub>;

50 mM  $\beta$ -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100  $\mu$ l 1 M  $\text{Na}_2\text{CO}_3$  gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK  
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im  
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

## Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Patterson, C., Goldstein, S., Younglai, E.V., Fletcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl. Acad. Sci.* 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). *Current protocols in molecular biology*. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. *Nature Genetics* 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. *Proc. Natl. Acad. Sci.* 92, 9082-9086.
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. *Science* 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. *Mol. Cell. Biol.* 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. *Science* 249, 1129-1133.
- 35 Harley, C.B., Fletcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. *Nature* 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with  
10 age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.  
15

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

**Patentansprüche**

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen  
Telomerase-Untereinheit.  
5
2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um  
Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,  
17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser  
Sequenzen handelt.  
10
3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um  
die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen  
katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder  
um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.  
15
4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der  
Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß  
es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide  
oder Proteine kodieren.  
20
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem  
der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.  
25
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vek-  
toren gemäß einem der Ansprüche 4 bis 6.  
30

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5           A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10
10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screenet.
- 15
11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20
12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25           mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

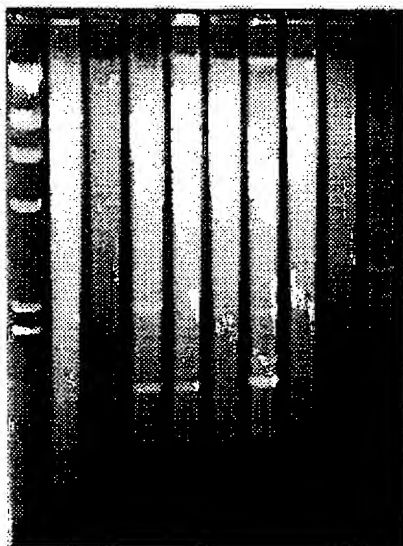
- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.



Fig. 1

A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10

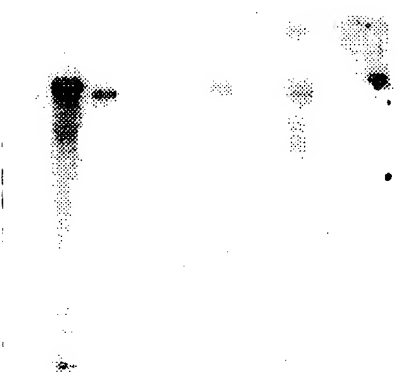


Fig. 2

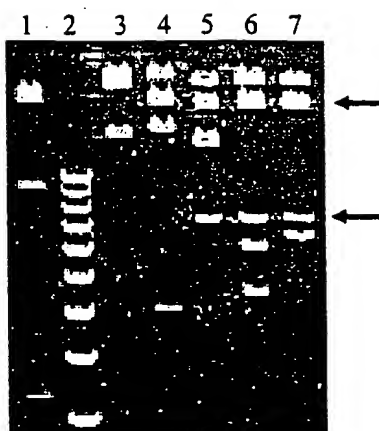


Fig. 3

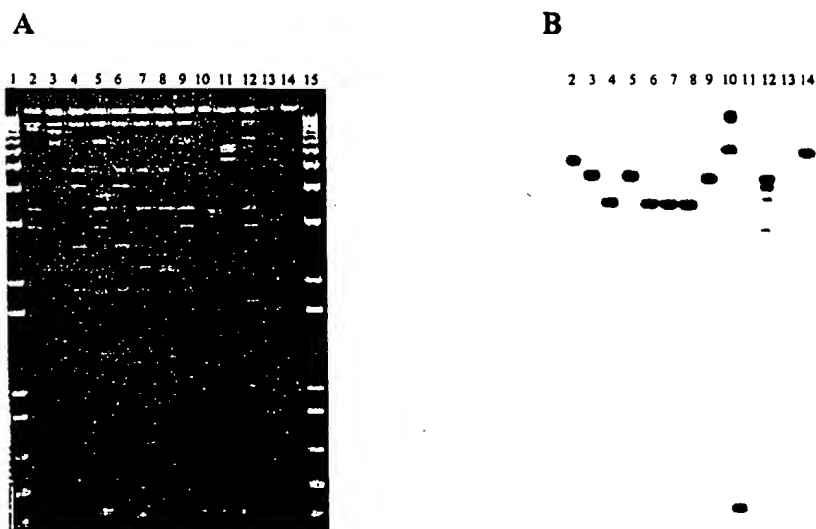




Fig. 4 (Fortsetzung)

```

CACAGCCTAG GCCGATTCTGA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGAAG CGGGGCCAG ACCCCCGGGT CCGCCCGGAG CAGCTCGCT GTCGGGGCA 4830
GGCCGGGCTC CCACTGGATT CGGGGCACA GACGCCAGG ACCGCGCTCC CCACGTGGCG GAGGACTGG 4900
GGACCCGGGC ACCCGTCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCGCCGTCCC 4970
GACCCCTCCC GGGTCCCCG CCCAGCCCC TCCGGGCCCT CCCAGCCCCT CCCCTTCCTT TCCGCGGCC 5040
CGCCCTCTCC TCGCGGCGG AGTTTCAGG AGCGCTGCGT CCTGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCC GCGATG 5126

```

Fig. 5

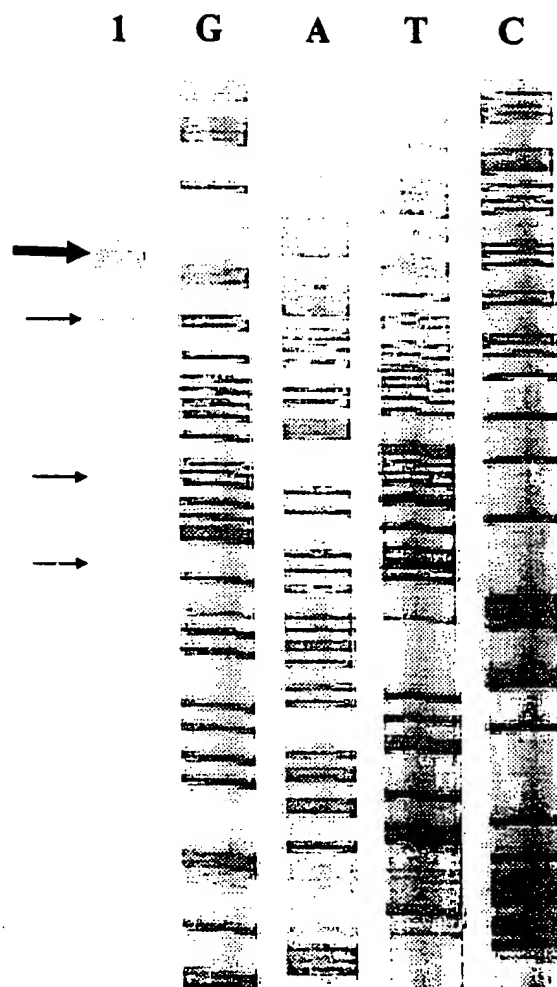




Fig. 7

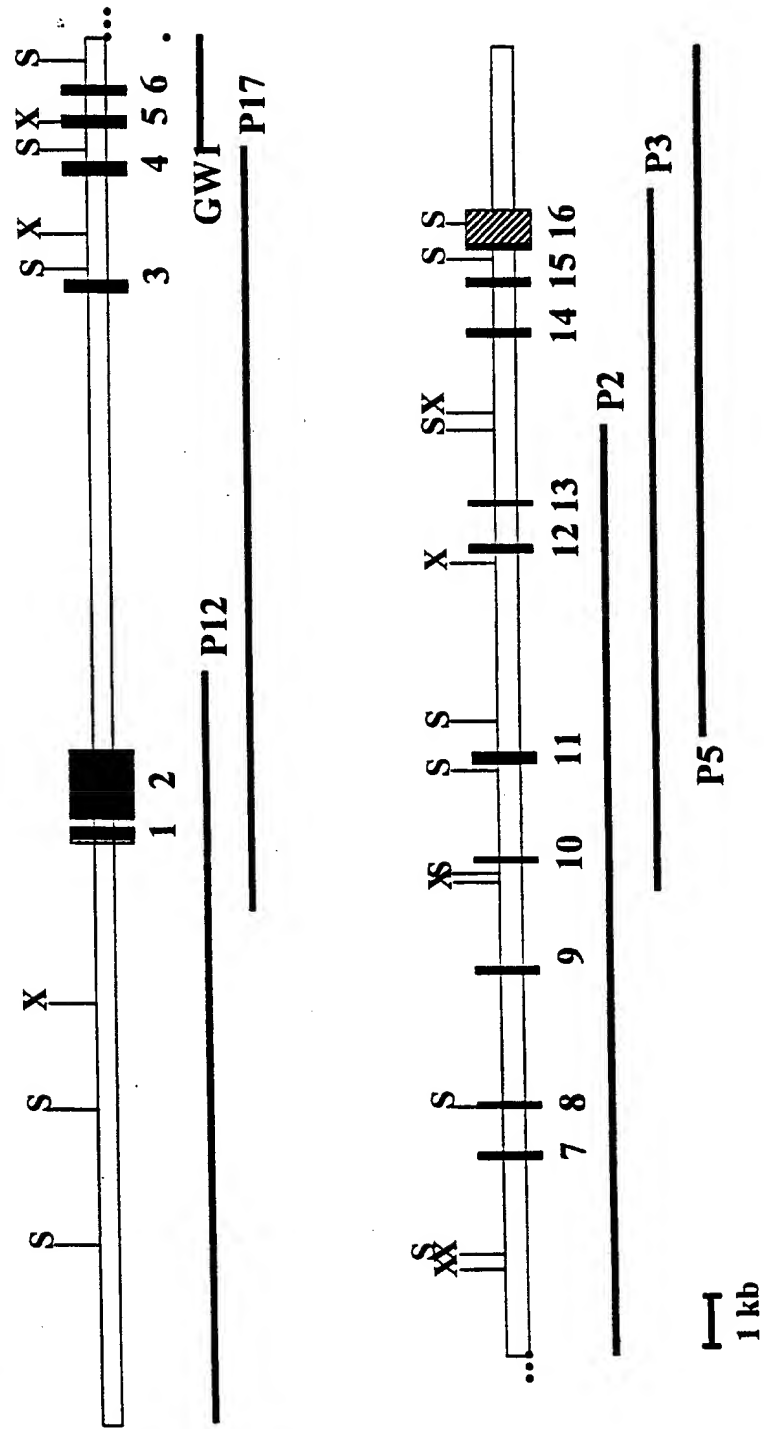




Fig. 8A

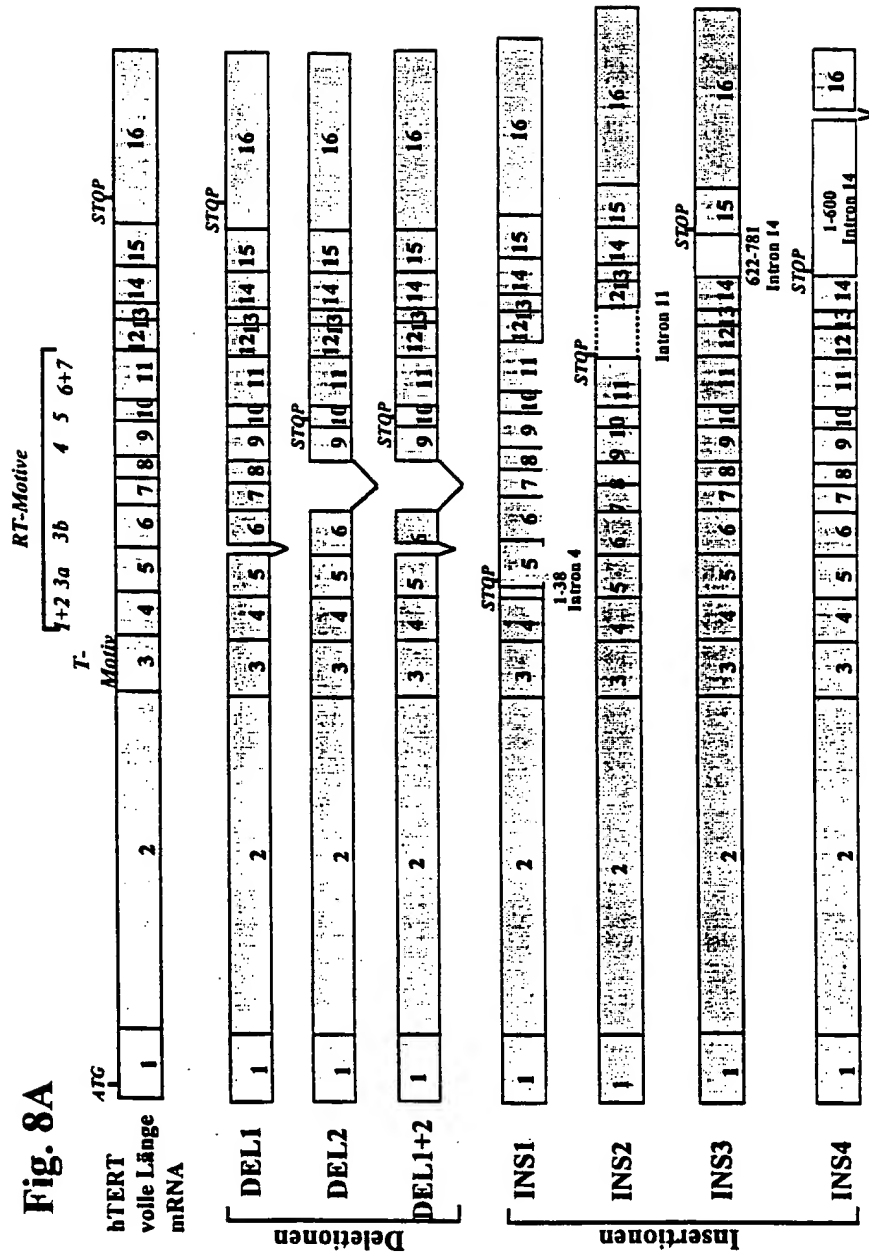
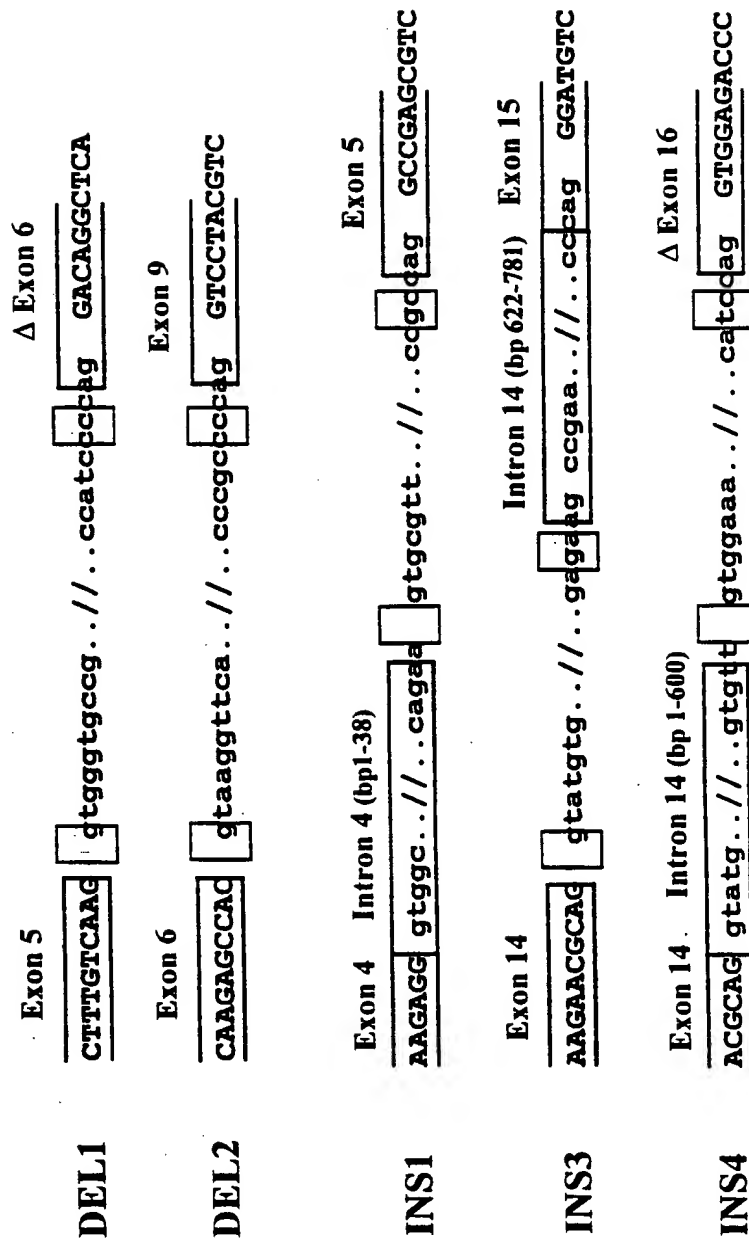


Fig. 8B



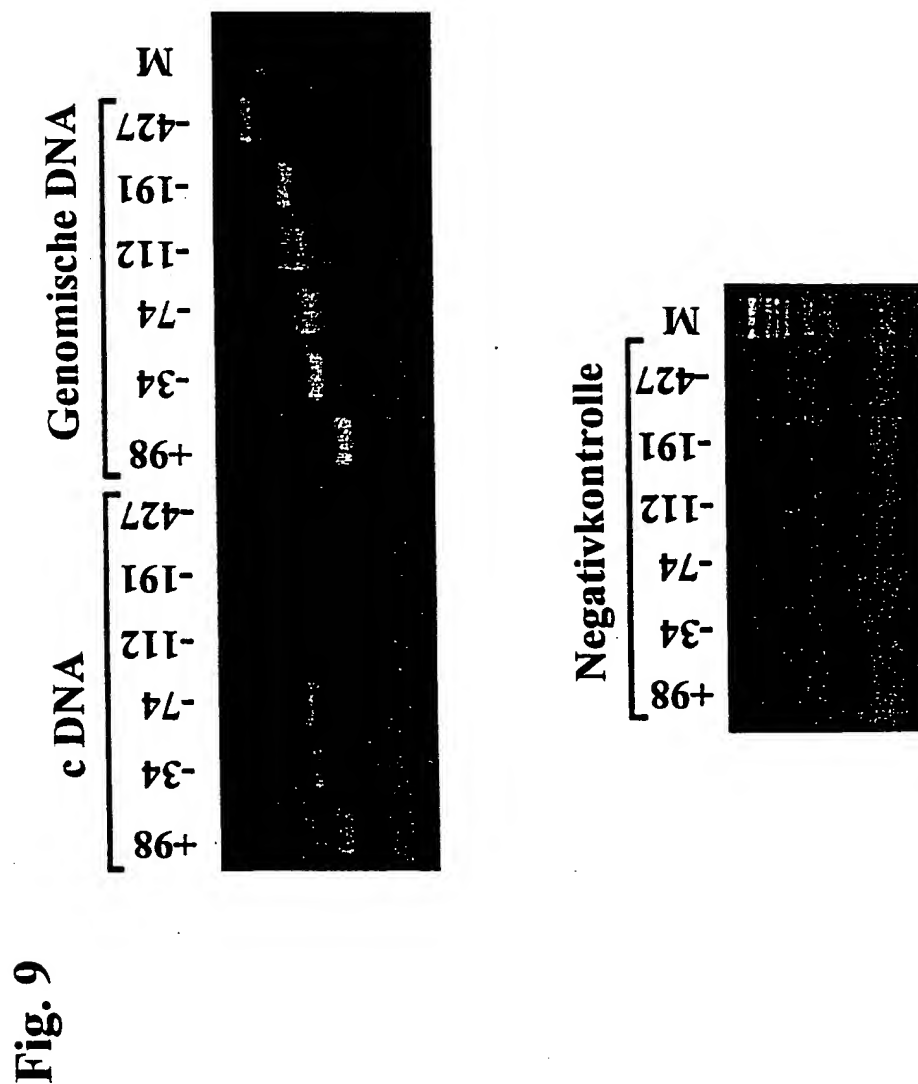


Fig. 10

ACTTGAGCCC AAGAGTTCAG GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204  
 ATGAGACCCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGGA -11134  
 AAAAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAAACT ATACAAACAC ATGAAAAATTA AACAAATATAC -11064  
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TAAAAAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA -10994  
 CGGAAACATA ACCTCTCAAA ACCCAGGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924  
 AGCAGCTACA TCAAAAAAGT AGAAAAAGCA GGCGCAGTGG CTCATGCTCG TAATCCCAGC ACTTTGGGAG -10854  
 GCCAAGGCGG GCAGATCGCC TGAGGTCAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCTG -10784  
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCAGATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG -10714  
 CGAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC -10644  
 TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAGAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT -10574  
 GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAAAGAA AGAAATAATA -10504  
 AAGATCAGAG CAGAAATAAA TGAACCTGAA AGATAACAA ACAAAAGATC AACAAAATTA AAAGTTGGTT -10434  
 TTTTGAAAAG ATAAACAAAA TTGACAAACC TTTGCCAGAG CTAAGAAAAA AGGAAAGAAG ACCTAAATAA -10364  
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA -10294  
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACAAC -10224  
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAAAAC GACCAATAAC AATAATGGGA TTAAGGCCAT -10154  
 AATAAAAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT -10084  
 AAAGAAGAAAT GAATTCCAAT CCTACTCAAA CTATTCTGAA AAATAGAGGA AAGAATACTT CCAAACTCAT -10014  
 TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAACA ACAAACAAAA -9944  
 CAGAAAGAAA GAAAACTACA GGCCAATATC CCTGATGAAT ACTGATACAA AAATCCTCAA CAAAACACTA -9874  
 GCAAAACAAA TTAACAACA CCTTCGAAAG ATCATTCTATT GTGATCAAGT GGGATTATT CCAGGGATGG -9804  
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAAACTA -9734  
 TATGATTATT TCACCTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTTCATGATA AAAACCTCTA -9664  
 AAAAACCCAGG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGGCATCCCC GCACTCTGGG -9594  
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGAG ACTAGCCTGG GCAACAAAAT GAGACCTGGT -9524  
 CTACAAAAAA CTTTTTTAAA AAATTAGCCA GGCGATGATG CATATGCCCTG TAGTCCCAGC TAGTCTGGAG -9454  
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCAGTGAGC CATGAACATG TCACCTGACT -9384  
 CCAGCCTAGA CAACAGAACA AGACCCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314  
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244  
 AAGAAACATA TTTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGCTGAA -9174  
 AGCCTTTCCT CTAAGATCTG GAAAAAGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104  
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAGGA AGAAGTCAAA -9034  
 TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGAAAA GACTTAAGAC ACCACTAAAA AACTATTAGA -8964  
 GCTGAAATTT GGTACAGCAG GATACAAAA CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC -8894  
 AAACAATCTG AAAAAGAAAC CAAAAAGCA GCTACAAATA AAATTAACA GCTAGGAATT AACCAAGAA -8824  
 GTGAAAGATC TCTACATGA AAATATAAAA ATGTTGATTA AAGAAATTGA AGAGGGCACA AAAAAAGAAA -8754  
 AGATATTCCA TGTTCTAGA TTGGAAGAAT AAATACTGTT AAAATGTCCA TACTACCCA AGCAATTTC -8684  
 AAATTCATG CAATCCCTAT TAAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTTCAAGAT -8614  
 TTGTACAGAA CCACAAAAGA CCCAGAATAG CCAAGCTAT CCTGACCAAA AAGAACAAA CTGGAAGCAT -8544  
 CACATTACCT GACTTCAAAT TATACTACAA AGCTATAGTA ACCCAAACTA CATGGTACTG GCATAAAAA -8474  
 AGATGAGACA TGGACCAGAG GAACAGAAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAACCTA -8404  
 TTTTGTACAA AGGTGCCAAG AACATACTTT GGGGAAAAGA TAATCTCTTC AATAAATGGT GCTGGAGGAA -8334  
 CTGGATATCC ATATGCAAAA TAACAATACT AGAAGCTGTG CTCTCACCAT ATACAAAAGC AAATCAAAAT -8264  
 GGTGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAAGAA AACACCGGAG AACCTCTCCA -8194  
 GGACATTGGA GTGGGCAAG ACTTCTTGAG TAATTCCCTG CAGGCACAGG CAACCAAGC AAAACAGAC -8124  
 AAATGGGATC ATATCAAGTT AAAAAGCTTC TGCCCAAGCA AGGAAACAAT CAACAAAGAG AAGAGACAAC -8054  
 CCACAGAATG GGAGAATATA TTTGCAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984  
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAGCTGAT TTTCAAAAAT AAGCAAAAGA TTTGGGTAGA -7914  
 CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA -7844  
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC -7774  
 AGGCAATTAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA -7704  
 TTGCTACCAC TATGGAGAAC AGTTTGAAG TTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC -7634  
 CCATTGCTAG GTATATACTC CAAAAAGGG AATCAGTGTA TCAACAAGCT ATCTCCACTC CCACATTTAC -7564  
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCAT CAACAGACGA ATGGAAAAG -7494  
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424  
 CAGCATGGGG GGCACTGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354  
 CTCCTTACT TGTGGGAGCA AAAATTAATA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG -7284  
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTATTG TATGTTTAA AATAACTAAA AGAGTATAAT -7214  
 TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCATTTTAC CCTGATGTGA -7144  
 TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCTTA CTATATTAAA -7074  
 AATTAAAAAT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCAGC ACTTTGGGAG GCCGAGGCGG -7004  
 GTGGATCACC TGAGGTCAGG AGTTTGAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTAATAAAGA -6934  
 TACAAAAATT AGCCAGGCGT GGTGGCAGAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864  
 TTGCTTGAAC CTGGGAGCGG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCAGTGCAGC CTGGGTGACA -6794  
 GAGCAAGACT CCATCTCAA AAAAAACAA AAAAAAGAAG ATTAATAATTG TAATTTTAT GTACCGTATA -6724  
 AATATATACT CTAATATATT AGAAGTTAAA AATTAACA ATTAATAAG GTAATTAACC ACTTAATCTA -6654  
 AAATAAGAAC AATGTATGTG GGGTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCACGA TGGCAGAAAT -6584

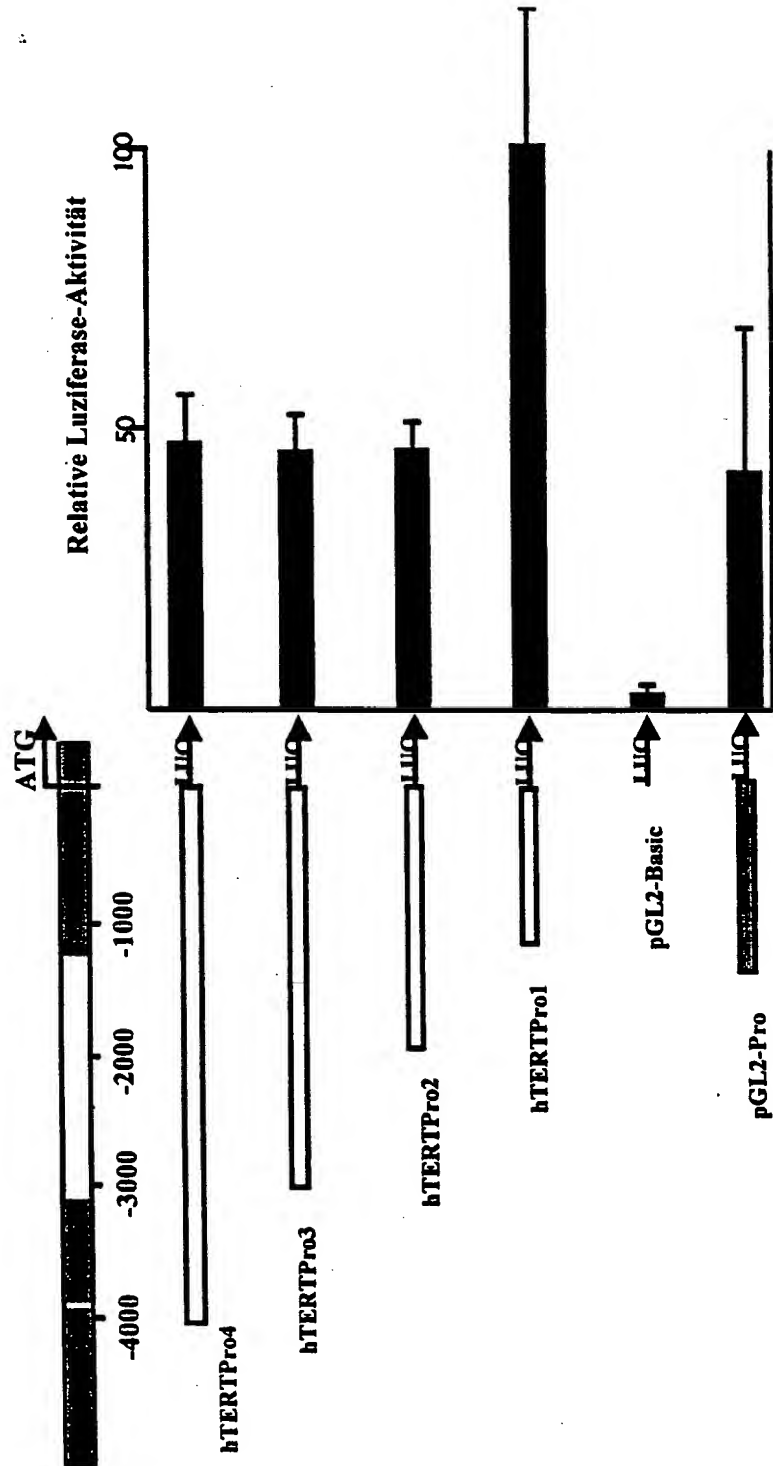
Fig. 10

GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAACC -6514  
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG -6444  
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374  
 CACCGTCTCT TCATTACCGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304  
 GTTAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAGAGAA TTAGAGTACC -6234  
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164  
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094  
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAG -6024  
 AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTCG TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954  
 CCCGGCCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA -5884  
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCCTCTT TAAACAGAA AGTCATGGAA -5814  
 GCACCTTCT CAAGGGAAGAA CCAGACGCCG GCTCTGCGGT CATTTACCTC TTTCTCTCTT CCCTCTCTTG -5744  
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCGTGGGA GCTTCTCCGA GCCCGTGTCT AGGACCTCT -5674  
 TGCAAGGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604  
 TGGCTGGGGG CGGACAGCGA CGCGGGGATT CAAAGACTTA ATTCCATGAG TAAATCAAC CTTTCCACAT -5534  
 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAATT TCATCAAATA ACATTCAGGA CTGCAGAAAT -5464  
 CCAAGGCGT AAAACAGGAA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394  
 ATTTTTCGCC CTAAGTACTT TTTATTGGTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAG GTACACGAGG -5324  
 AGAGGCCTGG GCGGCAGGGC TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254  
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGGCACA TTGGGCAACG CGAAGGCGGC CAGCTGCGT -5184  
 GTGACTCAGG ACCCCATACC GGCTTCTGG GCCCACCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114  
 CCGTGGAAAC GAACATGACC CTTGCCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA GTGGTGTGCA -5044  
 GGAAATGGCC ATGTAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATTTAT CATCTTACC -4974  
 CCAAGGACT GAATGATTC AGCAACTTCT TCGGGTGTGA CAAGCCATGA CAAACTCAG TACAAACACC -4904  
 ACTCTTTAC TAGGCCACA GAGCACGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834  
 GCTTTCAGCC ACCAGGCTGG GGTGACAACA GCGGTGAAC AGTCTGTTC TCTAGACTAG TAGACCTGG -4764  
 CAGGCACTCC CCCAGATTCT AGGCCTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694  
 CTGGGTGCC AACTGAGGC CAGCCCTGTC TCCACACCCT CCGCTCCAG GCCTCAGCTT CTCCAGCAGC -4624  
 TTCCTAAACC CTGGGTGGG CGTGTTCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGCG -4554  
 ACGTAGCTCG CACGGTTCCT CCTCACATGG GGTGTCTGTC TCCTTCCCCA AACTCACAT GCGTTGAAGG -4484  
 GAGGAGATTC TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGAGGTT -4414  
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCTC CTGTCTCTG CCGGGGCTG -4344  
 CCGGTGTGTT CTCTGTGTT TGCTGCTCCT TCCACGTCCA GCTGCGTGTG TCTCTGCCCG CTAGGCTCTC -4274  
 GGGGTTTTTA TAGGCATAGG ACGGGGGCGT GGTGGGCGAG GCGCTCTTG GGAATGCAA CATTGCGGTG -4204  
 TGAAAGTAGG AGTGCTGTC CTCACCTAGG TCCACGGGCA CAGGCCTGGG GATGGAGCCC CCGCCAGGGA -4134  
 CCCGCCCTTC TCTGCCAGC ACTTTCCTGC CCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064  
 ACTAAGCATC CTCTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATT TCCCCACAGC CCTGGGAATT -3994

c-Myc  
 CACGTGACTA CGCACATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924  
 CAAGCAGGG AAATCCCTGC TAAATGTCC TTTAACAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854  
 GGACAGTTTC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784  
 GAGTCAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTTT ACCCCATGGC -3714  
 AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCACGCA CCTTTTACTA -3644  
 AAGCGATGTT CTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574  
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGCAGG ATAATGCTCT AGAGATGCC -3504  
 ACGTCTGAT TCCCCAAAC CTGTGGACAG AACCCGCGCG GCCCAGGGC CTTTGCAGGT GTGATCTCCG -3434  
 TGAGGACCT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCCGAAA GTAATCCAGG GGTCTGGGA -3364  
 AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAAG -3294  
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCTCC -3224  
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCTCC -3154  
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA -3084  
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGCTGAA -3014  
 GTGCCTCCGG GCAAGGGCAG GGCAGGCACG AGTGATTTTA TTTAGCTATT TTATTTATT TACTTACTTT -2944  
 CTGAGACAGA GTTATGCTCT TGTGGCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874  
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTACAGC GTGCACCACC -2804  
 ACACCCGGCT AATTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAATC -2734  
 CTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664  
 GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594  
 AATTTCCCTT TTAATCAGGA GTTACCTCC TTTGATATT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524  
 CGTCTCTTGA CATATTACA GTTCTGTGA CCACCTGTTA TCCCATGGGA CCCACTGACG GGGCAGCTGG -2454  
 GAGGCTGCAG GCTTCAGGTC CCAAGTGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384  
 AAGTGTGGAC ACTGTCCTGA ATCTCAATGT CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314  
 CTCCTACTCT ACTGGGATTG AGCCCCCTCC CTATCCCCC CCAGGGGAG AGGAGTTCCT CTCACTCCTG -2244  
 TGGAGGAAGG AATGATACTT TGTATTGTT CACTGCTGGT ACTGAATCCA CTGTTTCACT TGTGGTTTG -2174  
 TTTGTTTTGT TTTGAGAGGC GGTTCACCTC TTGTGCTCA GGCTGGAGGG AGTGCAATGG CCGGATCTG -2104  
 GCTTACTGCA GCCTCTGCCT CCCAGGTTCA AGTGATTCTC CTGCTTCCGC CTCCCATTG GCTGGGATTA -2034  
 CAGGCACCCC CCACCATGCC CAGCTAATTT TTTGTATTT TAGTAGAGAC GGGGGTGGT GGGGTTACC -1964

ATGTTGGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894  
GATTACAGGT GTGAGCCACC ATGCCCAGCT CAGAATTAC TCTGTTTAGA AACATCTGGG TCTGAGGTAG -1824  
GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCCAA TGATAGAATT TTTTATTGT TGTTAGAACA -1754  
CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684  
ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCAATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614  
GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTTCCCT CTTTTAAAT TGTGTTTTCT -1544  
ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAAC TAACTTTGT TGGAACAAAT TTTCCAAACC -1474  
GCCCTTTGC CCTAGTGGCA GAGACAATTC ACAAACACAG CCCTTTAAAA AGGCTTAGGG ATCACTAAGG -1404  
GGATTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334  
GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAAGC AATTTCCTCC GGCAGTTTCT -1264  
GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTGTAGC ATTTCACTGT TTGCCGACCT CAGCTACAGC -1194  
ATCCCTGCAA GGCCTCGGGA GACCCAGAAG TTTCTCGCCC CTTAGATCC AAACCTGAGC AACCCGGAGT -1124  
CTGGATTCTT GGAAGTCCT CAGCTGTCCT GCGGTTGTGC CGGGGCCCA GGTCTGGAGG GGACCAGTGG -1054  
CGGTGTGGCT TCTACTGCTG GGCTGGAACT CGGGCCTCAAT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984  
GCTTGAGACC CGAGGCTGCC CTCACCCCTG TGCGGGCGGG ATGTGACCAG ATGTTGGCCT CATCTGCCAG -914  
ACAGAGTGCC GGGGCCAGG GTCAAGGCCG TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844  
GCGCCTGGCT CCATTCCCA CCGTTTCTCG ACGGGACCG CCGGCTGGGT GATTAACAGA TTTGGGGTGG -774  
TTTGCTCATG GTGGGGACCC CTCGCCGCT GAGAACCTGC AAAGAGAAAT GACGGGCTG TGTCAAGGAG -704  
CCCAAGTCGC GGGGAAGTGT TGCAGGGAG CACTCCGGGA GGTCCCGCT GCGCGTCCAG GGAGCAATGC -634  
GTCTCGGGT TCGTCCCGAC CGCGTCTAC GCGCTCCGT CCTCCCTTC ACGTCCGCA TTCGTGGTGC -564  
CCGGAGCCCG ACGCCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCCG GATCAGGCCA GCGGCCAAAG -494  
GGTCCCGCA CGCACCTGTT CCCAGGCCT CCACATCATG GCGGCTCCCT CGGGTTACCC CACAGCCTAG -424  
GCCGATTGCA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCT GGGAGCGCA GCGGCGCGG -354  
GGCGGGGAAG CGCGGCCAG ACCCCCCGGT CCGCGGGGAG CAGCTGCGCT GTCGGGGCCA GGCGGGGCTC -284  
CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CACGTGTCG GAGGGACTGG GGACCCGGGC -214  
ACCGTCTCTG CCCCTTACC TTCCAGCTCC GCCTCTCCG CGCGGACCC GCGGCTCCC GACCCCTCCC -144  
GGGTCCCCG CCCAGCCCC TCCGGGCCCT CCCAGCCCT CCCCTTCTCT TCCGCGGCC CGCCCTCTCC -74  
TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCG CACGTGGAAG CCCTGGCCCC GGCCACCCCC -4  
GCGATG

Fig.: 11



## SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen  
der humanen katalytischen Telomerase-Untereinheit und  
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

<140>  
<141>

15 <160> 20

<170> PatentIn Vers. 2.0

<210> 1  
<211> 5126  
20 <212> DNA  
<213> Homo sapiens

<400> 1  
25 gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60  
gggtaataa gtggtgtgca ggaatggcc atgtaaatta cagcactctg ctgatgggga 120  
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180  
tcgggtgtga caagccatga caaaactcag tacaaacacc actcttttac tagggccaca 240  
gagcacgggc cacacccctg atataatta agtccaggag agatgaggct gctttcagcc 300  
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360  
30 caggcactcc cccaaattct agggcctggg tgctgcttcc cgaggggcgc atctgcccctg 420  
gagactcagc ctgggggtgcc acactgaggc cagccctgtc tccacaccct ccgcctccag 480  
gccctcagctt ctccagcagc ttcttaaac ctgggtgggc cgtgttccag cgctactgtc 540  
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg caccgttctt cctcacatgg 600  
ggtgtctgtc tcttcccca acactcacat gcgtgaagg gaggagattc tgcgcctccc 660  
35 agactggctc ctctgagcct gaacctggct cgtggccccc gatgcagggt cctggcgctc 720  
ggctgcacgc tgacctccat ttccaggcgc tccccgtctc ctgtcatctg ccggggcctg 780  
ccggctgtgtt cttctgttcc tgtgtctcct tccacgtcca gctgcgtgtg tctctgccc 840  
ctagggtctc ggggttttta taggcatagg acgggggcgt ggtgggcccag ggcgctcttg 900  
ggaaatgcaa catttgggtg tgaaagtagg agtgcctgtc ctacactagg tccacgggca 960  
40 caggcctggg gatggagccc ccgccaggga cccgcccttc tctgccagc actttcctgc 1020  
ccccctccct ctggaacaca gagtggcagt ttccacaagc actaagcctc ctcttcccaa 1080  
aagaccacgc attggcacc ctggacattt gccccacagc cctgggaatt cactgacta 1140  
cgcacatcat gtacacactc ccgtccacga ccgacccccc ctgttttatt ttaatagcta 1200  
caaagcaggg aaatccctgc taaaatgtcc ttaacaaac tggttaaaca aacgggtcca 1260  
45 tccgcacggg ggacagtctc tcacagtga gaggaaacat ccgtttataa agcctgcagg 1320  
catctcaagg gaattacgct gagtcaaac tgccacctcc atgggatacg tacgcaacat 1380  
gctcaaaaag aaagaatttc accccatggc aggggagtg ttaggggggt taaggacggg 1440  
ggggggcgga gctgggggct actgcacgca ccttttacta aagccagttt cctggttctg 1500  
atggtatttg ctacgttatg ggagactaac cataggggag tggggatggg ggaacccgga 1560  
50 ggctgtgcca tctttgccat ccccgagtg cctgggcagg ataagtctct agagatgccc 1620  
acgtcctgat tcccccaaac ctgtggacag aaccgcgccg gccccagggc ctttgagggt 1680  
gtgatctccg tgaggacct gaggtctggg atccttcggg actacctgca ggcccgaaaa 1740  
gtaatccagg ggttctggga agaggcgggc agggagggtca gaggggggca gcctcaggac 1800  
gatggaggca gtcagtctga ggctgaaaag ggagggaggg cctcgagccc aggcctgcaa 1860  
55 gcgcctccag aagctggaaa aagcggggaa gggaccctcc acggagcctg cagcaggaa 1920  
gcacggctgg cccttagccc accaggggcc atcgtggacc tccggcctcc gtgccatagg 1980  
agggcactcg cgctgcccct ctacgatgaa gtgtgtgggg atttgagaa gcaacaggaa 2040  
acccatgcac tgtgaatcta ggattatttc aaaacaaagg ttacagaaa catccaagga 2100  
cagggtgtaa gtgcctccgg gcaagggcag ggcaggcacg agtgatttta ttagctatt 2160  
60 ttattttatt tacttacttt ctgagacaga gttatgctct tgttgcctcg gctggagtgc 2220  
tcagccctcc aagtagctgg gatttcaggc gtgcaccacc acacccggct aattttgtat 2340  
ttttagtaga gatgggcttt caccatgttg gtcaagctga tctcaaaatc ctgacctcag 2400  
gtgatccgcc caccctcagc tcccaaagtg ctgggattac aggcattgag cactgcacct 2460



2 / 18

5 ggccatttta accattttaaa aacttccctg ggctcaagtc acacccactg gtaaggagtt 2520  
 catggagttc aatttccctt ttactcagga gttaccctcc ttgtatattt tctgtaattc 2580  
 ttctgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640  
 tcccatggga cccactgcag gggcagctgg gaggtgcag gcttcaggtc ccagtggggt 2700  
 tgccatctgc cagtagaaac ctgatgtaga atcagggcgc aagtgtggac actgtcctga 2760  
 atctcaatgt ctcaagtgtg gctgaaacat gtgaaaatta aagtcacatc ctccactctc 2820  
 actgggattg agccccctcc ctatcccccc ccaggggcag aggagttcct ctactcctg 2880  
 tggagggaagg aatgatactt tgttattttt cactgctggt actgaatcca ctgtttcatt 2940  
 10 tgttgggttt tttgttttgt tttagagggc ggtttcactc ttgttgctca ggctggaggg 3000  
 agtgcaatgg cgcgatcttg gcttactgca gcctctgcct cccagggtca agtgattctc 3060  
 ctgcttccgc ctccatttg gctgggatta caggcaccgc ccaccatgcc cagctaattt 3120  
 tttgtatttt tagtagagac ggggttgggt ggggttcacc atgttggtcca ggctgggtcc 3180  
 gaacttctga cctcagatga tccactgcc tctgcctcct aaagtgcctg gattacaggt 3240  
 gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300  
 15 gaagctcacc ccaactcaagt gttgtgggtg ttttaagccaa tgatagaatt tttttattgt 3360  
 tgttagaaca ctcttgatgt ttacactgt gatgactaag acatcatcag tttttcaaag 3420  
 acacactaac tgcaccata atactgggt gtcttctggg tatcagcaat ctctattgaa 3480  
 tgccgggagg cgtttctctg ccatgcacat ggtgttaatt actccagcat aatcttctgc 3540  
 ttccatttct tctcttccct cttttaaaat tgtgttttct atgttggtct ctctgcagac 3600  
 20 aaccagtgtg agctacaact taacttttgt tggaaacaaat ttccaaacc gccctttg 3660  
 cctagtggca gagacaattc acaaacacag ccttttaaaa aggcttaggg atcactaagg 3720  
 ggatttctag aagagcgacc tgaatccta agtatttaca agacgaggtc aacctccagc 3780  
 gagcgtgaca gccaggggag ggtgcgaggg ctgttcaaat gctagctcca taaataaagg 3840  
 aatttctctc ggcagtttct gaaagtagga aaggttacat ttaagggtgc gtttcttagc 3900  
 25 atttcagtgt ttgccgacct cagctacagc atccctgcaa ggcctcggga gaccagaag 3960  
 tttctcgccc ccttagatcc aaacttgagc aacctggagt ctggattcct gggaagtcct 4020  
 cagctgtcct gcggttctgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080  
 tctactgctg ggtcggaagt cgggctcctc agctctgcag tccgaggctt ggagccaggt 4140  
 gcctggagccc cgaggctgcc ctccaccctg tgcgggcggg atgtgaccag atgttgacct 4200  
 30 catctgccag acagagtgcc gggggccagg gtcaaggccg ttgtggctgg tgtgagggc 4260  
 ccggtgcgag gccagcagga gcgcctggct ccatttccca ccttttctcg acgggaccgc 4320  
 cccggtgggt gattaacaga tttgggtggt tttgctcatg gtgggggacc ctgcgcgctc 4380  
 gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtcgc ggggaagtgt 4440  
 tgcagggagg cactccggga ggtcccgcgt gcccgccag ggagcaatgc gtcctcgggt 4500  
 35 tgcgtccccc cgcgctctac gcgcctccgt cctcccttcc acgtccggca ttcgtgggtg 4560  
 ccggagcccg acgccccgcg tccggacctg gaggcagccc tgggtctccg gatcaggcca 4620  
 gcggccaaaag ggtcgccgca cgcacctgtt cccagggcct ccacatcatg gccctcctc 4680  
 cgggttacct cacagcctag gccgattcga cctctctccg ctggggccct cgtggcgctc 4740  
 cctgcacctt gggagcgcga gcgcgcgcgc ggcgggggag cgcggcccag acccccgggt 4800  
 40 ccgcccggag cagctgcgct gtcggggcca ggcggggctc ccagtggatt cgcggggcaca 4860  
 gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggaccggggc acccgctctg 4920  
 ccccttcacc ttccagctcc gcctcctccg cgcggacccc gccccgctcc gacccctccc 4980  
 ggggtccccg cccagccccc tccgggcccct cccagcccct ccccttctct tccgcgccc 5040  
 45 cgccctctcc tccggggcgc agtttcaagg agcgtgcgt cctgctgcgc acgtgggaaag 5100  
 ccctggcccc ggccaccccc gcgatg 5126

&lt;210&gt; 2

&lt;211&gt; 4042

&lt;212&gt; DNA

50 &lt;213&gt; Homo sapiens

&lt;400&gt; 2

55 gtttcaggca gcgctgcgct ctgctgcgca cgtgggaagc cctggccccc gccacccccg 60  
 cgatgccgcg cgctccccgc tgcgagccg tgcgctccct gctgcgcagc cactaccgcg 120  
 aggtgctgcc gctggccacg ttctgtgcggc gcctggggcc ccagggtctg cggtcgtg 180  
 agcgcgggga cccggcggtt ttccgcgcgc tggtagccca gtgcctggtg tgcgtgccct 240  
 gggacgcacg gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300  
 tggtagcccc agtgctgcag aggtctgtcg agcgcggcgc gaagaacgtg ctggccttgc 360  
 60 gcttcgcgct gctggacggg gcccgccggg gccccccga ggccttcacc accagcgtgc 420  
 gcagctacct gcccacacg gtgaccgacg cactgcgggg gagcggggcg tgggggctgc 480  
 tgcgtgcgcg cgtggggcag gacgtgtgg ttcacctgct ggcacgctgc gcgctcttgc 540  
 tgcgtgtggc tccagctgc gcctaccagg tgtgcggggc gccgctgtac cagctcggcg 600  
 ctgccactca gggcgggccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgcg 660  
 aacgggcctg gaaccatagc gtcagggagg ccgggggtccc cctgggacct ccagccccgg 720  
 65 gtgcgaggag gcgcgggggc agtgccagcc gaagtctgcc gttgcccagg aggccccagg 780

gtggcgctgc ccctgagccg gagcggacgc ccgttgggca ggggtcctgg gcccaccg 840  
 gcaggacgcg tggaccgagt gaccgtgggt tctgtgtggt gtacactgcc agaccgcg 900  
 aagaagccac ctctttggag ggtgcgctct ctggcacgcg cactccccc ccatccgtgg 960  
 gcccgcagca ccacgcgggc ccccatcca catcgcgcc accacgtccc tgggacacgc 1020  
 5 cttgtccccc ggtgtacgcc gagaccaagc acttcctcta ctctcagge gacaaggagc 1080  
 agctgcggcc ctcccttcta ctacgtcttc tgaggcccg cctgactggc gctcggaggc 1140  
 tcgtggagac catctttctg ggttccaggc cctggatgcc agggactccc cgaggttg 1200  
 cccgcctgcc ccagcgctac tggcaaatgc ggccctgtt tctggagctg cttgggaacc 1260  
 acgcgcagtg cccctacggg gtgctcctca agacgcactg cccgctcgca gctcgggtca 1320  
 10 ccccagcagc cgggtgtctg gcccgggaga agccccagg ctctgtggcg gccccgagg 1380  
 aggaggacac agacccccgt cgcctgtgag agctgctccg ccagcacagc agccccggc 1440  
 aggtgtacgg ctctgtgagg gcttcctgc gccggctggt gcccacaggc ctctggggct 1500  
 ccaggcaca cgaacggcgc ttccctcagg acaccaagaa gttcatctcc ctggggaagc 1560  
 atggcaagct ctgcctgacg gagctgacgt ggaagatgag cgtgcgggac tgcgcttg 1620  
 15 cctgcaggag cccaggggtt ggctgtgttc cggccgcaga gcaccgtctg cgtgaggaga 1680  
 tcttgcccaa gtctctgac tggtgtatga gtgtgtactg cgtcagactg ctgaggtctt 1740  
 tcttttatgt caggagacc acgtttctaa agaacaggct ctttttctac cggagagtg 1800  
 cctggagcaa gttgcaagc attggaatca gacagcactt gaagagggtg cagctgcggg 1860  
 agctgtcgga agcagaggtc aggcagatc gggaagccag gccgcctctg ctgacgtcca 1920  
 20 gactccgctt catccccaag cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980  
 tgggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg aggggaaagg 2040  
 cactgttcag cgtgctcaac tacgagcggg cgcggcgccc cggcctctg ggcgctctg 2100  
 tgctgggccc ggacgatc caccgggccc ggcgcacctt cgtgctcgct gtgcgggccc 2160  
 aggacccgccc gcctgagctg tactttgtca aggtggatgt gacgggcccg tacgacacca 2220  
 25 tccccagga caggctcacg gaggtcatcg ccagcatcat caaaccccag aacacgtact 2280  
 gcgtgcgtcg gtatgcccgt gtcagaaagg ccgcccattg gcacgtccgc aaggccttca 2340  
 agagccacgt ctctaccttg acagacctcc agccgtatag gcgacagttc gtggctcacc 2400  
 tgcaggagac cagcccgtg agggatgccc tcgtcatcga gcagagctcc tccctgaatg 2460  
 30 aggccagcag tggcctcttc gacgtcttcc tacgcttcat gtgccaccac gccgtgcgca 2520  
 tcaggggcaa gtccctacgt cagtgcagg ggcctccgca gggctccatc ctctccacgc 2580  
 tgctctgcag cctgtgctac ggcgacatgg agaacaagct gtttgccggg attcggcggg 2640  
 acgggctgct cctgcgtttg gtggatgatt tctgttgggt gacacctcac ctccaccag 2700  
 cgaatacctt ctccaggacc ctgggtccgag gtgtccctga gtatggctgc gtggtgaact 2760  
 35 tgcggaagac agtgggtaac ttccctgtag aagacgaggc cctgggtggc aggcttttg 2820  
 ttcagatgcc ggcccacggc ctattccctt ggtgcggcct gctgctggat acccgagacc 2880  
 tggaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcact 2940  
 tcaaccgagg ctccaaggct gggaggaaca tgcgtcgcaa actctttggg gtcttgcggc 3000  
 tgaagtgtca cagcctgttt ctggatttgc aggtgaacag cctccagacg gtgtgacca 3060  
 acatctacaa gatcctcctg ctgcaggcgt acaggtttca cgcagtgtg ctgcagctcc 3120  
 40 catttcatca gcaagtttgg aagaacccca catttttctt gcgcgtcatc tctgacacgg 3180  
 cctccctctg ctactccatc ctgaaagcca agaacgcagg gatgtcgtg ggggccaagg 3240  
 gcgcgcggc cctctgccc tccgagggcg tgacgtggct gtgccaccaa gcattctctg 3300  
 tcaagctgac tcgacacgt gtcaactacg tgccactcct ggggtcactc aggacagccc 3360  
 45 agacgcagct gactcggaag ctcccgggga cgacgtgac tgccctggag gccgcagcca 3420  
 acccggcact gccctcagac ttcaagacca tccctggactg atggccaccc gccacagcc 3480  
 aggcggagag cagacaccag cagccctgtc acgcccggct ctacgtccca gggaggagg 3540  
 ggcggccccc acccagggc gcaccgctgg gactctgagg cctgagtgag tgtttggccg 3600  
 aggcctgcat gtccggctga aggtgagtg tccggctgag gcctgagcga gtgtccagcc 3660  
 aagggtgtag tgtccagcac acctgcccgt ttcacttccc cacaggctgg cgctcggctc 3720  
 50 caccacaggg ccagcttttc ctaccagga gcccggcttc cactccccac ataggaatag 3780  
 tccatcccca gattcgccat tgttaccccc tcgcccgtgc ctcttttgc ttccaccccc 3840  
 accatccagg tggagacct gagaaggacc ctgggagctc tgggaaattg gactgaccaa 3900  
 aggtgtgccc tgtacacagg cgaggacct gcacctgag ggggtccct gtgggtcaaa 3960  
 55 ttggggggag gtgctgtggg agtaaaatag tgaatatag agtttttcag ttttgaaaaa 4020  
 aaaaaaaaaa aaaaaaaaaa aa 4042

<210> 3  
 <211> 11276  
 <212> DNA  
 60 <213> Homo sapiens

<400> 3  
 acttgagccc aagagttcaa ggctacggtg agccatgatt gcaacaccac acgcccagct 60  
 tgggtgacaga atgagacctt gttcaaaaa aaaaaaaaaa aattgaaata atataaagca 120  
 65 tcttctcttg ccacagtggg acaaaaccag aaatcaacaa caagagggaat tttgaaact 180

atacaaacac atgaaaatta aacaatatac ttctgaatga ccagtgaagc aatgaagaaa 240  
 ttaaaaagga aattgaaaaa tttatTTtaag caaatgataa cggaaacata acctctcaaa 300  
 acccacggta tacagcaaaa gcagtgtctaa gaaggaaagt tatagctata agcagctaca 360  
 tcaaaaaagt agaaaaagcca ggcgcagtgg ctcatgcctg taatcccagc actttgggag 420  
 5 gccaaggcgg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480  
 aaccttctcg ctactaaaaa taaaaaatta gctgggcctg gtggcacatg cctgtaatcc 540  
 cagctactcg ggaggctgag gcaggataac cgtttgaacc caggaggtgg aggttgcggt 600  
 gagccgggat tgcgccattg gactccagcc tgggtaacaa gagtgaacc ctgtctcaag 660  
 10 aaaaaaaaaa aagtagaaaa acttaaaaat acaacctaat gatgcacctt aaagaactag 720  
 aaaaagcaaga gcaaaactaaa cctaaaaattg gtaaaagaaa agaaataata aagatcagag 780  
 cagaaataaa tgaactgaa agataacaat acaaaagatc acaaaaatta aaagtTgggt 840  
 ttttgaaaag ataaacaaaa ttgacaaacc tttgccaga ctaagaaaaa aggaagaag 900  
 acctaaataa ataaagttag agatgaaaaa agagacatta caactgatac cacagaaatt 960  
 caaaggatca ctagaggcta ctatgagcaa ctgtacacta ataaatrgaa aaacctgaa 1020  
 15 aaaaatagata aattcctaga tgcatacaac ctaccaagt tgaaccatga agaaatccaa 1080  
 agcccaaaac gaccaataac aataatggga ttaaagccat aataaaaagt ctcttagcaa 1140  
 agagaagccc aggacccaat ggcttccctg ctggatttta ccaatcattt aaagaagaat 1200  
 gaattccaat cctactcaaa ctattctgaa aaatagagga aagaatactt ccaaactcat 1260  
 20 tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaaacaaac 1320  
 aacaaaaaaa cagaagaaa gaaaactaca ggccaatatc cctgatgaat actgatacaa 1380  
 aaatcctcaa caaaactca gaaaacaaa ttaaaacaa ccttcgaaag atcattcatt 1440  
 gtgatcaagt gggatttatt ccagggatgg aaggatgggt caacatagc aaatcaatca 1500  
 atgtgataca tcatcccaac aaaatgaagt acaaaaaacta tatgattatt tcactttatg 1560  
 cagaaaaagc atttgataaa attctgcacc ctctcatgata aaaacctca aaaaaccagg 1620  
 25 tatacaagaa acatacaggc caggcacagt ggctcacacc tgcgatccca gcactctggg 1680  
 aggcccaagg gggatgattg cttgggcccc gaagtTtgag actagcctgg gcaacaaaat 1740  
 gagacctggc ctacaaaaa cttttttaa aaattagcca ggcgatgagg catatgcctg 1800  
 tagtcccagc tagtctggag gctgagggtg gagaatcact taagcctagg aggtcgaggc 1860  
 30 tgcagtgagc catgaacatg tcaactgtact ccagcctaga caacagaaca agacccact 1920  
 gaataagaag aaggagaagg agaagggaga agggaggagg aaggaggagg gaggagaag 1980  
 agggaggTgga ggagaagtgg aagggggaagg ggaagggaag gaggaagaag aagaacata 2040  
 tttcaacata ataaaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100  
 agcctttcct ctaagatctg gaaaatgaca agggccact ttcaccactg tgattcaaca 2160  
 35 tagtactaga agtccctagt agagcaatca gataagagaa agaaataaaa ggcatccaaa 2220  
 ctggaaaagg agaagtcaaa ttatcctgtt tgcagatgat atgatcttat atctggaaaa 2280  
 gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaa 2340  
 caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaaatctg aaaaagaaa 2400  
 caaaaaagca gctacaaata aaattaaaca gctaggaatt aaccaaagaa gtgaaagatc 2460  
 40 tctacaatga aaactataaa atgttgataa aagaaattga agagggcaca aaaaagaaa 2520  
 agatattcca tgttcataga ttggaagaat aaatactgtt aaaaatgtcca tactcccaa 2580  
 agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaaat 2640  
 agaagaaaca attctaagat ttgtacagaa ccacaaaaga ccagaaatag ccaaaagctat 2700  
 cctgaccaa aagaacaaa ctggaagcat cacattacct gacttcaaat tatactacaa 2760  
 45 agctatagta acccaaaacta catggtagtg gcataaaaac agatgagaca tggaccagaa 2820  
 gaacagaata gagaatccag aaacaaatcc atgcatctac agtgaactca tttttgacaa 2880  
 aggtgccaag aacatacttt ggggaaaaa taatctcttc aataaatggt gctggaggaa 2940  
 ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcccat atacaaaagc 3000  
 aaatcaaaaat ggatgaaagg cttaaatcta aaacctcaa ctttgcaact actaaaagaa 3060  
 50 aacaccggag aaactctcca ggacattgga gtgggcaaa acttcttgag taattccctg 3120  
 caggcacagg caaccaaaagc aaaaacagac aaatgggagc atatcaagt aaaaagcttc 3180  
 tggccagcaa aggaacaaat caacaaagag aagagacaac ccacagaatg ggagaatata 3240  
 tttgcaaaact attcatctaa caaggaatta ataaccagta tatataagga gctcaaaacta 3300  
 ctctataaga aaaacaccta atagctgat tttcaaaaat aagcaaaaga tctgggtaga 3360  
 catTTtcaa aataagtcat acaaatggca aacaggcatc tgaaaatgtg ctcaacacca 3420  
 55 ctgatcatca gagaatgca aatcaaaact actatgagag atcatctcat cccagttaaa 3480  
 atggctttta ttcaaaagac aggaataaac aaatgccagt gaggatgtgg ataaaaggaa 3540  
 acccttgagc actgttggtg ggaatggaaa ttgctaccac tatggagaac agtttgaaag 3600  
 ttcctcaaaa aactaaaaat aaagctacca tacagcaatc ccattgtctag gtatacactc 3660  
 caaaaaaggg aatcagtgtg tcaacaagct atctccactc ccacatttac tgcagcactg 3720  
 60 ttcatagcag ccaaggtttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780  
 aaaatgtggg gcacatacac aatggagtac tacgcagcca taaaaagaa tgagatcctg 3840  
 tcagtTgcaa cagcatgggg ggcactgggc agtatgttaa gtgaaataag ccaggcacag 3900  
 aaagacaaac ttttcatgtt ctcccttact tgtgggagca aaatttaaaa caattgacat 3960  
 65 agaaatagag gagaatgggt gttctagagg ggtgggggac aggggtgacta gagtcaacaa 4020  
 taatttattg tatgttttaa aataactaaa agagtataat tgggtgtgtt gtaacacaaa 4080

5 / 18

5 gaaaggataa atgcttgaag gtgacagata ccccatTTac cctgatgtga ttattacaca 4140  
 ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200  
 aattaaaaatt ttaatggcca ggcacggctg ctcattgtccg taatcccagc accttgggag 4260  
 gccgaggcgg gtggatcacc tgaggctcagg agtttgaaac cagtctggcc accatgatga 4320  
 aacctgtctt ctactaaaga tacaaaaatt agccaggcgt ggtggcacat acctgtagt 4380  
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgag 4440  
 tgagccgaga tcatgccact gcaactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500  
 acaaaaaaaa aaaaaagaag attaaaaatt taatttttat gtaccgtata aatatatact 4560  
 ctactatatt agaagttaaa aattaaaaca attataaaag gtaattaaacc acttaactta 4620  
 10 aaataagaac aatgtatgtg gggtttctag cttctgaaga agtaaaagt atggccacga 4680  
 tggcagaaat gtgaggagg aacagtggaa gttactgtt ttagacgctc atactctctg 4740  
 taagtgtact aattttaacc aaagacaggc tgggagaagt taaagaggca ttctataagc 4800  
 cctaaaaaaa ctgctaataa tggtgaaagg taatctctat taattaccaa taattacaga 4860  
 tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtcttc tcattcacgg 4920  
 15 tgcctttttt cttgtgtgct tggagatttt cgattgtgtg ttcgtgtttg gttaaactta 4980  
 atctgtatga atcttgaaac gaaaaatggt ggtgatttcc tccagaagaa ttagagtacc 5040  
 tggcaggaag caggtggctc tggggacctg agccacttca atcttcaagg gtctctggcc 5100  
 aagaccagg tgcaaggcag aggcctgatg acccgaggac aggaaagctc ggtgggag 5160  
 gggcgatgag aagcctgctc cgttggtag cagcgcatga agtgcctta ttacgctt 5220  
 20 gcaagatgt ctctggatc catctgaaa agcgccagc cgggaatgca aggagtcaga 5280  
 agcctctctc tcaaacccag gccagcagct atggcgccca cccggcgctg tgccagagg 5340  
 agaggagtca aggcacctc aagtatgct taaatctttt ttccacctga agcagtgc 5400  
 aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcatgga 5460  
 gcaccctctc caagggaaaa ccagacgccc gctctcggt cattacctc ttctctctc 5520  
 25 cctctctctg cctctcggt tctgtatcgg gacagagtga ccccgctgga gcttctcga 5580  
 gcccgctcgt aggacctct tgcaaaaggc tccacagacc cccgctctg agagaggagt 5640  
 ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcggaat 5700  
 caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttgatttta 5760  
 tcttaataat ttcttaaat tcaatcaata acattcagga ctgcagaaat ccaaggcgt 5820  
 30 aaaaacaggaa ctgagctatg tttgccaagg tccaaggact taataacat gttcagagg 5880  
 atttttcgcc ctaagtactt tttattggtt ttcataagg gtcttaggt gcaaggga 5940  
 gtacacgagg agaggcctg gggcagggc tatgagcagc gcagggccac cggggagaga 6000  
 gtcccccggc tgggaggctg acagcaggac cactgacct cctccctgg agctgccaca 6060  
 35 ttgggcaacg cgaaggcggc cagcctgctg gtgactcagg acccataacc ggttctctg 6120  
 gcccaaccac actaacccag gaagtcaagg agctctgaac ccgtggaaac gaacatgacc 6180  
 cttgcctgcc tgcctccctg ggtgggtcaa gggtaatgaa gtggtgtgca ggaatggcc 6240  
 atgtaaaata cagcactctg ctgatgggga cegtctctc catcattatt catcttacc 6300  
 cccaaggact gaatgattcc agcaacttct tgggtgtga caagccatga caaaactcag 6360  
 tacaacacac actcttttac tagggccaca gacacggsc cacaccctg atatatag 6420  
 40 agtcaggagg agatgagct gctttcagcc accaggctgg ggtgacaaca ggggtgaa 6480  
 agtctgttcc tctagactag tagacctgg caggcactcc ccagattct agggcctgg 6540  
 tgctgcttcc cgaggggccc atctgccctg gagactcagc ctgggtgtcc acactgagg 6600  
 cagcctgtc tccacacct cgcctccag gctcagctt ctccagcagc ttcttaacc 6660  
 45 ctgggtgggc cgtgttccag cgctactgtc tcacctgtcc cactgtgtct tgtctcagc 6720  
 acgtagctcg cagggttct cctcacatgg ggtgtctgtc tcttcccca acactcacat 6780  
 gcgttgaagg gaggagattc tgcgctccc agactggctc ctctgagcct gaacctggct 6840  
 cgtggccccc gatgcaggtt cctggcgtcc ggtgcacgc tgacctccat ttccaggcgc 6900  
 tccccgtctc ctgtcatctg cgggggctg cgggtgtgtt ctctgtttc tgtgtctct 6960  
 50 tccacgtcca gctgcgtgtg tctctgccc ctagggtctc ggggttttta taggcatagg 7020  
 acggggcggt ggtggggcag ggcgtcttg ggaatgcaa catttgggtg tgaaagtagg 7080  
 agtgctctgc ctacactagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140  
 cccgcccctc tctgcccagc actttcctgc cccctccct ctggaacaca gaggcagc 7200  
 ttccacaagc actaagcatc ctcttcccaa aagaccagc attgccaccc ctggacattt 7260  
 55 gcccacagc cctgggaatt cagtgacta cgcacatcat gtacacact cgttccaca 7320  
 ccgacccccg ctgttttatt ttaatagcta caaagcagg aaatccctgc taaaatgtcc 7380  
 ttaacaaaac tggttaaaca aacgggtcca tccgacggg ggacagttcc tcacagtga 7440  
 gaggaacatg ccgtttataa agcctgcagg catctcaagg gaattacgct gactcaaaa 7500  
 tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc accccatggc 7560  
 60 aggggagtggt ttaggggggt taaggacggt gggggcgga gctgggggt actgcacgca 7620  
 ccttttacta aagccagttt cctggtctg atggtattg ctgattatg ggagactaac 7680  
 cataggggag tggggatggg ggaaccggga ggtgtgcca tctttgccat gcccgagtgt 7740  
 cctgggcagg ataagtctct agagatgccc acgtcctgat tccccaaac ctgtggacag 7800  
 aaccgccccg gcccagggc ctttgcaggt gtgactctcg taggaccct gaggtctggg 7860  
 65 atccttcggg actacctgca ggcccgaaaa gtaatccagg ggttctggga agaggcgggc 7920  
 agggagggtca ggggggggca gcctcaggac gatggaggca gtcagctcga ggctgaaa 7980

6 / 18

5 ggagggaggg cctcgagccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040  
 gggaccctcc acggagcctg cagcaggaa gacggctgg cccttagccc accagggccc 8100  
 atcgtggacc tccggcctcc gtgcatagg agggcactcg cgctgccctt ctgcatgaa 8160  
 gtgtgtgggg atttgagaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220  
 5 aaaaacaaagg ttacagaaa catccaagga cagggtgaa gtgcctccgg gcaagggcag 8280  
 ggcaggcacg agtgatttta tttagctatt ttattttatt tacttacttt ctgagacaga 8340  
 gttatgctct tgtgcccag gctggagtgc agcggcatga tcttggtca ctgcaacctc 8400  
 cgctcctctg gttcaagcaa ttctcgtgcc tcagcctccc aagtagcttg gatttcaggc 8460  
 10 gtgcaccacc acaccggctt aattttgtat tttagtaga gatgggcttt caccatgttg 8520  
 gtcaagctga tctcaaatc ctgacctcag gtgatccgcc caccctagcc tcccaaatg 8580  
 ctgggattac aggcagtagc cactgcacct ggccatttta accatttta aacttccctg 8640  
 ggctcaagtc acaccactg gtaaggagtt catggagttc aatttccctt ttactcagga 8700  
 gttaccctcc ttgtatattt tctgtaattc ttctgtagct ggggatacac cgtctcttga 8760  
 15 catattcaca gtttctgtga ccacctgtta tcccatggga cccactgcag gggcagcttg 8820  
 gaggctgcag gcttcaggtc ccagtggggg tgccatctgc cagtagaaac ctgatgata 8880  
 atcagggcgc aagtgtggac actgtcctga atctcaatgt ctgagtgtgt gctgaaacat 8940  
 tagaaaatta aagtcacatc ctctactctt actgggattg agcccccttc ctatcccccc 9000  
 ccaggggcag aggagtctct ctactcctg tggaggaaag aatgatactt tgtatttttt 9060  
 cactgctggg actgaatcca ctgtttcatt tgttggtttg ttgtttttgt ttgagaggc 9120  
 20 ggtttcactc ttgttgctca ggctggaggg agtgcaatgg cgcgactctg gcttacttga 9180  
 gcctctgctt cccagggtca agtgattctc ctgcttccgc ctcccatttg gctgggatta 9240  
 caggcacccg ccaccatgcc cagctaattt ttgtatttt tagtagagac gggggtgggt 9300  
 ggggttcacc atgttgcca ggctggtctc gaacttctga cctcagatga tccacctgcc 9360  
 tctgctctct aaagtgtgg gattacaggt gtgagccacc atgccagct cagaatttac 9420  
 25 tctgtttaga aacatctggg tctgaggtag gaagctcacc ccactcaagt gttgtggtgt 9480  
 ttttaagcaa tgatagaatt tttttattgt tgttagaaca ctcttgatgt ttacactgt 9540  
 gatgactaag acatcatcag cttttcaag acacactaac tgcaccata atactgggtg 9600  
 gtcttctggg tatcagcaat cttcattgaa tgccggagg cgtttctctg ccactgacat 9660  
 30 ggtgttaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaat 9720  
 tgtgttttct atgttggtt ctctgcagag aaccagtgtg agctacaact taacttttgt 9780  
 tggaaacaaat tttccaaacc gcccctttgc cctagtggca gagacaattc acaaacacag 9840  
 ccccttaaaa aggcttaggg atcactaagg ggatttctag aagagcgacc tgtaatccta 9900  
 agtatttaca agacgaggct aacctccagc gagcgtgaca gcccagggag ggtgcgaggc 9960  
 35 ctgttcaaat gctagctcca taaataaagc aatttccctc ggcagtttct gaaagttaga 10020  
 aaggttcatat ttaagggtgc gtttgttagc atttcagtgt ttgccagct cagctacagc 10080  
 atccctgcaa ggctcggga gaccagaag tttctcgcct ccttagatcc aaacttgagc 10140  
 aaccggagt ctggattcct gggaagtcct cagctgtcct gcggttgtgc cggggcccca 10200  
 ggtctggagg ggaccagtgg ccgtgtggct tctactgctg ggctggaaat cgggcctcct 10260  
 40 agctctgcag tccgaggctt ggagccaggt gcctggacc cagaggtgct ctcaccctg 10320  
 tgcggggggg atgtgaccag atgttggtc catctgccag acagagtgc gggggccagg 10380  
 gtcaaggccg ttgtggctgg tgtgaggcgc ccggtgcgcg gccagcagga gcgcctggct 10440  
 ccatttccca ccttctctcg acgggacccg cccgggtgggt gattaaacaga ttgggggtg 10500  
 tttgtctatg gtggggaccg ctgcgcctc gagaacctgc aaagagaaat gacgggctg 10560  
 45 tgtcaaggag cccaagtgc ggggaagtgt tgcagggagg cactccggga ggtcccgcgt 10620  
 gcccgctccag ggagcaatgc gtccctcgggt tgcctccag ccgcgtctac gcgcctcctg 10680  
 cctccctctc acgtccggca ttctgtgtgc ccggagcccg acgccccgcg tccggacctg 10740  
 gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgccgca cgcacctgtt 10800  
 cccagggcct ccacatcatg gcccctccct cgggttacc caccgctag gccgattcga 10860  
 50 cctctctccg ctggggccct cgctggcgtc cctgcacctt gggagcgca gccggcgcg 10920  
 ggcggggaag cgcggccag accccgggtt ccgcccggag cagctgcgt gtccggggcca 10980  
 ggccgggtct ccagtggatt cgcgggcaca gacggccagg acccgctcc ccactgtggt 11040  
 gagggactgg ggaccgggg acccgtctct ccccttacc ttccagctcc gcctcctccg 11100  
 cgcggacccc gccccgtccc gaacctccc ggggtcccg cccagccccc tccgggcccct 11160  
 55 cccagccctt ccccttctct tccgggccc cgcctctctc tcgcgcgcg agtttcaggc 11220  
 agcgctgcgt cctgctgcgc acgtgggaag ccctggcccc ggcaccccc gcgatg 11276

<210> 4  
 <211> 104  
 <212> DNA  
 60 <213> Homo sapiens

<400> 4  
 gtgggcctcc ccggggctcg cgtccggctg ggggttaggg cggccggggg gaaccagcga 60  
 catgaggaga gcagcgcagg cgactcagg cgcttcccc gcag 104

<210> 5  
<211> 8616  
<212> DNA  
<213> Homo sapiens

5

<400> 5

gtgaggaggt ggtggccgtc gagggcccag gccccagagc tgaatgcagt aggggctcag 60  
aaaagggggc aggcagagcc ctggctcctc tgtctccatc gtcacgtggg cacacgtggc 120  
ttttcgtctc ggacgtcgag tggacacggg gatctctgcc tctgctctcc ctctgtcca 180  
10 gtttcgataa acttacgagg ttccacctca cgttttgatg gacacgaggg ttccaggcgc 240  
cgaggccaga gcagtgaaca gagggaggctg ggcgcggcag tggagccggg ttgcccggca 300  
tggggagaag tgtctggaag cacagacgct ctggcgaggg tgcctgcagg ttacctataa 360  
tctctctcgc aatttcaagg gtgggaatga gaggtgggga cgagaacccc ctcttctctg 420  
gggtgggagg taagggtttt gcaggtgcac gtggtcagcc aatatgcagg tttgtgttta 480  
15 agatttaatt gtgtgttgac gggcagggtg ggtggctcac gccggtaatc ccagcacttt 540  
gggaagctga ggcaggtgga tcacctgagg tcaggagttt gagaccagcc tgaccaacat 600  
ggtgaaaccc tatctgtact aaaaatacaa aaattagctg ggcaggtgtg tgtgtgcctg 660  
taatcccagc tacttgggag gctgaggcag gagaatcact tgaacccagg aggcggaggc 720  
tgcagtgagc tgagattgtt ccattgtact ccagcctggg cgacaagagt gaaactcctg 780  
20 ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagatagg 840  
actgttctcc agcacagatc ctggctcccat ctttaggtat gaagagggcc acatgggagc 900  
agaggacagc agatggctcc acctgtctgag gaaggagacg tgtttgtggg tgttcagggg 960  
atgggtctgc tgggccctgc cgtgtcccca cctgtttttt ctggatttga tgttgaggaa 1020  
cctccgctcc agcccccttt tggctcccag tgcctccagg cctaccctg gcagctagaa 1080  
25 gaagtcccca ttccaccccc tccccacaaa ctcccaagac atgtaagact tccggccatg 1140  
cagacaagga ggggtgacct cttggggctc tttttttctt tttttttctt ttatgggtgg 1200  
aaaagtcata taacatgaga ttggcactcc taacaccgtt tctgtgtgac agtgcagaa 1260  
tgctaactcg gcggtgttta cagcagggtt cttgaaatgc tgcgtcttgc gtgactggaa 1320  
gtccctaccc atcgaaaggc agctgcctca cactgtctgc ggctcagggt gaccacggcg 1380  
30 agtcagataa gcgtcatgac acccagtttt gctttttgtg ctccagcttc cttcgttgag 1440  
gagagtttga gttctctgat caggactctg cctgtcattg ctgttctctg acttcagatg 1500  
aggtcacaaat ctgccccctg cttatgcagg gaggtagggc tgggtcccggt ggtgtccctg 1560  
cacgtgcagg gtgagtgagg cgttggcccc aggtgtccct gtcacgtgta gggtagtgta 1620  
35 ggcgcggccc ccgggtgtcc ctgtcccggt cagcgtgatt gagggtgtggc ccccggtgtg 1680  
ccctgtcacg ttaggggtga gtgaggcgcc atccccgggt gtccctgtca cgtgtagggt 1740  
gagtgaggcg tgggtcccggt gtgtccctgt cccgtgcagg gtgagtgagg cactgtcccc 1800  
gggtgtccct gtcacgtgca ggtgagtgga ggcgcgggtc ccgggtgtcc ctctcaggtg 1860  
taggggtgag gaggcgcgcc ccaggggtgt cctgttcacg ttaggggtga gtgaggcacc 1920  
gtccctgggt gtccctccca ggtatagggt gaggtaggca ctgtcccggt ggtgtccctg 1980  
40 cacgtgcagg gtgagtgagg cgcggccccc ggggtgtccct ctacagggtga gggtagtgta 2040  
ggcgtctgcc ctgggtgtcc ctgtctctgt taggggtgag gaggctctgt cccaggtgtg 2100  
ccttggcggt tgcctcactg agcttgcctc tgaatgtttg ctctttctat agccacagct 2160  
gcgcgggttg cccattgcct gggtagatgg tgcaggcgca gtgctgggtcc ccaagcctat 2220  
cttttctgat gctcggctct tcttggctac ctctccgttc cattttgcta cggggacagc 2280  
45 ggactgcagg ctctcgcttc ccgcgtgcca ggcactgcag ccacagcttc aggtccgctt 2340  
gcctctgttg ggcctggctt gctcaccacg tgcggccac atgcatgctg ccaataactcc 2400  
tctccagctt tgtctcatgc cgaggctgga ctctgggctg cctgtgtctg ctgccacgtg 2460  
ttgttgga catcccagaa agggttctct gtgccctgaa ggaaagcaag tcaccccagc 2520  
50 cccctcactt gtcctgtttt ctcccaagct gccctctgct tgggccccct tgggtgggtg 2580  
gcaacgcttg tcaccttatt ctgggcacct gccgtcatt gcttaggctg ggtctgtgct 2640  
ccagtcgccc cctcacatgg attgacgtcc agccacagg tggagtgctt ctgtctgtct 2700  
cctgtctga gaccacgtg gaggggcggg gtctccgcca gccttcgtca gacttccctc 2760  
ttgggtctta gttttgaatt tcactgattt acctctgacg ttcttatctc tccattgtat 2820  
gctttttctt ggtttattct ttcattcctt tcttagcttc ttagtttagt catgctcttc 2880  
55 cctctaagtg ctgccttacc tgcacctgt gttttgatgt gaagtaatct caacatcagc 2940  
cactttcaag tgttcttaaa atacttcaaa gtgttaatac ttcttttaag tattcttatt 3000  
ctgtgatttt tttctttgtg cagctgtgtt tttagcgtga aatcattttg atatcagtga 3060  
cttttaagta ttcttttagc tattctgtga ttcttttagc cagtgagttt ttgaaact 3120  
60 gtttatgttc aagatatgta gagtatcaag atacgtagag tatttttaag tatcatttta 3180  
ttattgattt ctaactcagt tgtgtagtgg tctgtataat accaattatt tgaagtttgc 3240  
ggagccttgc tttgtgatct agtgtgtgca tggtttccag aactgtccat tgtaaatttg 3300  
acatcctgtc aatagtgggc atgcatgttc actatatcca gcttattaag gtccagtcga 3360  
aagcttctgt ctcttctag atgcatgaaa ttccaagaag gaggccatag tccctcacct 3420  
65 gggggatggg tctgttcatt tcttctcgtt tggtagcatt tatgtgaggc attgttaggt 3480  
gcacgcacgt ggtagaattt ttacttctct gatgagtga tcttttgag acttctatgt 3540

ctctagtaat ctagtaattc tttttttaaa ttgctcttag tactgccaca ctgggcttct 3600  
 tttgattagt attttccctgc tgtgtctgtt ttctgccttt aatttatata tatatatata 3660  
 tttttttttt ttttgagaca gagtcttggt ctgtcgccca ggggtgagtgc agtgggtgta 3720  
 5 tcacaggcca gtgtaacttt taccttcttg cctgagccgt cctctcacct cagcctcctg 3780  
 agtagctgga actgcagaca cgcaccgcta cacttggtta atttttaaat tttttctgga 3840  
 gacagggtct tgcctgtgtg cccaggctgg tctcaaaactc ttggactcaa gggatccatc 3900  
 tacctcggct tcccaaagt ctgaattaca ggcatgagcc accatgtctg gcctaatttt 3960  
 caacactttt atattcttat agtgtgggta tgcctgtta acagcatgta ggtgaatttc 4020  
 caatccagtc tgacagtcgt tgtttaactg gataacctga ttatttttca tttttttgtc 4080  
 10 actagagacc cgctcgtgct actctgattc tccacttgcc tggttgcatgt cctcgttccc 4140  
 ttgttttcca ccactcttg ggttgccatg tgcgtttcct gccgagtgtg tgttgatcct 4200  
 ctgctgctcct cctggctcact gggcatttgc ttttatttct ttttgcctag tgttaccctc 4260  
 tgatcttttt attgtcgttg tttgcttttg tttattgaga cagtctcact ctgtcaccca 4320  
 ggtctggagt taatggcaca atctcggctc actgcaacct ctgcctctc ggttcaagca 4380  
 15 gttctcattc ctcaacctca tgagtgcgtg ggattacagg cggccaccac cagcctcctg 4440  
 taatttttgt atttttagta gagataggct ttcaccatgt tggccaggct ggtctcaaac 4500  
 tcttgacctc aagtgatctg cccgccttgg cctccacacag tgcctgggatt acaggtgcaa 4560  
 gccaccgtgc ccggcatacc ttgatctttt aaaaagaagt ctgaaacatt gctacccctg 4620  
 20 tcttgagcaa taagacctt agtgtatttt agctctggcc acccccagc ctgtgtgtgt 4680  
 ttttccctgc tgacttagtt ctactcagg catcttgaca ccccacacag ctaagcatca 4740  
 ttaattattgt tttccgtgtc gagtgtttct gtacttttgc ccccgcctg cttttctccc 4800  
 tttgttcccc gctgtgtctt tgtctcaggc ccgctgctg ggggtccctt ctttgcctt 4860  
 tgcgtgggtc tctgtcttg ttattgctgg taaaccccag ctttacctgt gctggcctcc 4920  
 atggcatcta ggcagctccg gggacctctg cttatgatgc acagatgaag atgtggagac 4980  
 25 tcacgaggag ggcggtcacc ttggccctg agtgtctgga gcaccacgtg gccagcgttc 5040  
 cttagccagt gagtgcacgc aacgtccgct cggcctgggt tcagcctgga aaaccccagg 5100  
 catgtcgggg tctggtggct ccgctggtgc gagtgtgaaa tcgctgcaaac ctgctggtgtg 5160  
 gcgccagctc tgacggtgct gctcggggg gtagtgtctg cttctctcct tctgcttggg 5220  
 aaccagaca aaggatgagg ctccgagccg ttgtcgcaca acaggagcat gacgtctgga 5280  
 30 atgtggataa ttttaaaatt tctaggctgg gcgctggtgc tcacgcctgt aatcccagca 5340  
 ctttgggagg ccaaggcggg tggatcacga ggtcaggagg tcgagaccat cctggccaac 5400  
 atgatgaac cccatctgta ctaaaaacac aaaaattagc tgggctggtt ggcgggtgcc 5460  
 tgtaatccca gctactcggg aggtcaggc aggagaattg cttgaacctg ggagctggaa 5520  
 35 gttgcagtga gccgacattg caccactgca ctccagcctg gcaacacagc gagactctgt 5580  
 ctcaaaaaaa aaaaaaaa aaaaaaaa aattctagta gccacattaa aaaagttaaa 5640  
 aagaaaagggt gaaattaatg taataataga ttttactgaa gccacagcat tccacacctc 5700  
 atcatttttag ggtgttatgt gtgggagcat cactcacagg acatttgaca ttttttgagc 5760  
 tttgtctgcg ggtacccgtg tgtaggctcc gtgctggtgc atctcggcct ggacctgctg 5820  
 40 ggcttcccat ggcatggct gttgtaccag atgggtgcagg tccgggatga ggtcgcagg 5880  
 ccctcagtga gctggatgct cagtgtccgg atgggtgcag tctgggatga ggtcgcagg 5940  
 ccctgctgtg agctggatgt gtggtgtctg gatgtgcag gtcagggtgt aggtctccag 6000  
 gccctcgggt agctggaggt atggagctcg gatgatgcag gtcgggggtg aggtcgccag 6060  
 gccctgctgt gagctggatg tgtggtgtct ggtggtgca ggtcagggtt gaggctctca 6120  
 45 ggcctcgggt aagctggagg tatggagctc ggatgatgca ggtcgggggt gaggctgcca 6180  
 ggcctgctg tgagctggat gtgtggtgtc tggatggtgc aggtctgggt tgaggtcacc 6240  
 aggcctcgtg gtgagctggg tgtgctggtc ctggatggtg caggctctgga gtgaggtcgc 6300  
 cagacggtgc cagacctgct ggtgagctgg atatcggtg tccggatggt gcaggtctg 6360  
 ggtgaggttg ccaggccctg ctgtgagttg gatgtgggt gtcgggatgc tgcaggtccg 6420  
 50 gtgtgaggtc accaggccct gctgtgagct ggtgtgtggt tgtctggatg gtgaggtct 6480  
 ggggtgaagg tccgccaggcc cctgcttgtg agctggatgt gttgtgtctg gatggtgca 6540  
 gtctggagtg aggtcgccag gccctcgggt agctggatgt gcaggttcca gatggtgca 6600  
 gtccgggtgt aggtcgccag accctgctgt gagctggatg tgcgtgtgtt ggtggtgca 6660  
 ggtctggagt gaggctcgcca gccctcgggt gagctggatg tatggagctc ggtggtgcc 6720  
 55 ggtccgggtg gaggctcgcca gacctgctg tgagctggat gtcgggtgtc tggatggtac 6780  
 aggtctggag tgaggtcgcc agacctgct gtgagctgga tatgctggtt ccggatggtg 6840  
 caggctcagg gtgaggtctc caggccctcg gtgagctgga ggtatggagt ccggatgatg 6900  
 caggctcggg gtgaggtcgc caggccctgc tgtgaactgg atgtgcggcg tctggatggt 6960  
 gcaggtctgg ggtgtggtcg ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020  
 60 gcaggtccgg ggtgaggtcg ccaggccctg ctgtgagctg gatgtgcggc gtcctgatgg 7080  
 tgcaggtctg ggtgtgtgtc gccaggccct cgtgtgagct gaggatgga gtcgggatga 7140  
 tgcaggtccg ggtgtgaggt gccaggccct gctgtgagct ggatgtgctg tatccggatg 7200  
 gtgcagctcg ggtgtgaggt gccaggccct gctgtgagct ggatgtgctg tatccggatg 7260  
 gtgcaggtct ggtgtgaggt caccaggccc tgcgtgagc tggatgtgctg ggtcgggt 7320  
 65 gctgcaggtc cgggtgaggt tccaggccc ctcggtgagc tggatgtgctg ggtcgggt 7380  
 gtccggatgg tgcaggtcca ggtgtgaggt gctaggccct tgggtgggtg gatgtgctg 7440

5  
10  
15  
20

gtccggatgg tgcaggtctg gggtagagtc gccaggccct tggtagagctg gatgtgcggg 7500  
gtctgcatgg tgcaggtctg gggtagagtc gccaggccct tggtagagctg gatgtgtggg 7560  
gtccggatgg tgcaggtccg gctgtagagtc gccaggccct gctgtgagct ggtgtgctgg 7620  
tgctctggatg gtgcaggtcc ggggtgaggt agccaaggcc ttcggtgagc tggatgtggg 7680  
gtgtccggat ggtgcaggtc cggggtgaggt tgcgccaggcc ctgctggttag ctggatatgc 7740  
gggtgtccgga tgggtcaggt cggggtgaggt gtcaccaggc cctgctggtta gctggatgtg 7800  
cgggtgtctgg atggtgcaggt tccggggtga ggtcgcaggc cctgctgtgt agctggatgt 7860  
gctgtatccg gatggtgcag gtccggggtg aggtcgcagg gccctgcagt gactgtggatg 7920  
tgctgtatcc ggatggtgca ggtctggtg gagggtcgcca ggccctgctg ttagctggat 7980  
atgcggtgtc ggatggtgca ggtccggggt gagggtcacca ggccctgctg ttagctggat 8040  
gtgcggtgtc cggatggtgc aggtctggtg tgaggtcgcc aggcctgtgt gtgagctgga 8100  
tggtctgtat ccggtggtg cagggtccggg gtgaggtcgc caggccctgc ggtgagctgg 8160  
atgtgctgta tccggatggt gcaggtctgg cgtgaggtcg ccaggccctg cgggtgagctg 8220  
gatgtgcagt gtacggatgg tgcaggtccg ggtgtaggtc gccaggccct gctggtgggt 8280  
gtatgtgtgt tgtctggatg gtgcaggtcc ggggtgaggt cgccaggccc tgcggtgagc 8340  
tggtatgtgt gtgtctggat gtcgaggtc cggggtgaggt tgcgccaggcc ctggtgagc 8400  
tggatatgct gtgtcccggt gtccgaatgg tgcaggtcca gggtagaggt gccaggccct 8460  
tggtagagctg gatgtgcgtg gtccggatgg tgcaggtctg gggtagaggt gccaggccct 8520  
tggtagagctg gatgtgcgtg gtccggatgg tgcaggtccg gggtagaggt accaggccct 8580  
cggtagatctg gatgtggcat gtccctctctg ttttaag 8616

&lt;210&gt; 6

&lt;211&gt; 2089

&lt;212&gt; DNA

25 &lt;213&gt; Homo sapiens

&lt;400&gt; 6

30  
35  
40  
45  
50  
55  
60

gtactgtatc cccacgccag gcctctgctt ctggaagtc tggaaacacca gcccggccct 60  
agcatgcccc tgtctccact tgctctgctt tccctggctg tgcagctctg ggtggggagc 120  
caggggcccc gtccacaggcc tgggtccaaag gattctctgt caaggctctg actgctctga 180  
gctcacgttc tcttacttgt aaaatcagga gtttgtgcca agtggctctt aggggtttga 240  
aagcagaagg gatttaaaat agatggaaac actaccacta gcctccttgc ctttcccttg 300  
gatgtgggtc tgattctctc tctctttttt ttttctttt tgagatggag tctcactctg 360  
ttgccaggc tggagtgtag tggcataatc ttggctcact gcaacctcca cctcctgggt 420  
ttaagcgatt caccagcctc agcctcctaa gtactgtgga ttacaggcac ctgccaccac 480  
gcctgggttaa tttttgtact ttttaggagag acgggggttc accatgtgtg ccaggctggg 540  
ctcgaaactc tgacctcagg tgatccaccc accttgccct cccaaagtgc tgggtttaca 600  
ggctaagcca cgtgccccag cccccgattc tcttttaatt catgctgttc tgtatgaatc 660  
ttcaatctat tggatttagg tcatgagagg ataaaatccc acccacttgg cgaactcactg 720  
caggagacac ctgtgcaggg agcacctggg gataggagag ttccaccatg agctaacttc 780  
taggtggctg catttgaatg gctgtgagat ttgtcttcca atgttcggct gatgagagtg 840  
tgagattgtg acagattcaa gctggatttg catcagttag ggacgggagc gctggtcttg 900  
gagatgccag cctggctgag cccaggccat ggtattagct tctccgtgtc ccgcccaggc 960  
tgactgtgga gggctttagt cagaagatca gggcttcccc agctccccctg cacactcgag 1020  
tccctggggg gccttgtgac acccctgccc ccaaatcagg atgtctgcag agggagctgg 1080  
cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccagac tgggtgctgg 1140  
gccatttctt tgcattctgg ggagggctcag ggcctttccct gtgggaacaa gttataacac 1200  
aatgcacctt acctagactt tacacgtatt taatgggtgt cgacccaaca tggctcattg 1260  
accagttatt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtggtggg 1320  
ccccaaagat ctccctgtga ctactgggac tgttgttctg cctggggggg cttggaggcc 1380  
cctcctccct ggacagggta cgtgccttt tctactctgc tgggctgctg gcctgcggtg 1440  
agggcaccag ctccggagca cccgcggccc cagtgtccac ggagtgccag gctgtcagcc 1500  
acagatgccc aggtccaggt gtggccgctc cagccccctg gcccccagtg gtggttttgg 1560  
gggaaaaggc caagggcaga ggtgtcagga gactgggtgg ctcatgagag ctgattctgc 1620  
tccttggctg agctgccctg agcagcctct cccgccctct ccactggaag ggtgtggct 1680  
ctttctacct ggggttctg cctggggcca gccttgggt accccagtgg ctgtaccaga 1740  
gggacaggca tccctgtgtg aggggcatgg gttcacgtgg cccagatgc agcctgggac 1800  
caggctccct ggtgtgatg gtgggacagt caccctgggg gttgaccgac ggaactggcg 1860  
tccccagggt tgactatagg accaggtgtc cagggtgcct gcaagtagag gggctctcag 1920  
aggcgtctgg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgtgctg 1980  
tgggtgccct gagccctcac tgagtcggtg ggggcttctg gcttcccctg agcttcccc 2040  
tagtctgtg tctggctgag caagcctcct gaggggctct ctattgcag 2089



10 / 18

5  
5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

<210> 7  
<211> 687  
<212> DNA  
<213> Homo sapiens

<400> 7  
gtggctgtgc tttggtttaa cttccttttt aaacagaagt gcgtttgagc cccacatttg 60  
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120  
cggcgccaac ccattttgtg gcacagttag gtggccgagg tggcgggtgcc tccagaaaag 180  
cagcgtgggg gtgtaggggg agctcctggg gcagggacag gctctgagga ccacaagaag 240  
cagccggggc agggcctgga tgcagcacgg cccgaggtcc tggatccgtg tcctgctgtg 300  
gtgcgcagcc tccgtgcgct tccgcttacg gggcccgggg accaggccac gactgccagg 360  
agcccaccgg gctctgagga tcctggacct tgccccacgg ctctctgcacc ccacccctgt 420  
ggctgcggtg gctgcggtga ccccgctatc tgaggagagt gtggggtgag gtggacagag 480  
gtgtggcatg aggatcccgt gtgcaacaca catgcggcca ggaacccgtt tcaaacaggg 540  
tctgaggaaag ctgggagggg ttctagggtc cgggtctggg tggctgggga cactggggag 600  
gggtctgttc tcccctgggt ccctatgggt ggggtgggac ttggccggat ccactttcct 660  
gactgtctcc catgctgtcc ccgccag 687

<210> 8  
<211> 494  
<212> DNA  
<213> Homo sapiens

<400> 8  
gtgggtgccc gggacccccg tgagcagccc tgctggacct tgggagtggc tgccgtattg 60  
gcacctcatg ttgggtggag gaggtactcc tgggtgggccc gcaggagtg caggtgaccc 120  
tgtactgtt gaggcacacac ctggcaccta ggttgaggcc ctccagccct tcctgcagca 180  
catggggccc actgtgcacc ctgactgccc gggctcctat tcccaggag ggtccactg 240  
gattccagt tccgtcagag aaggaaaccg aacggctcag ccaccaggcc ccggtgcctt 300  
gcaccccagt cctgagccag gggctcctg tcctgaggct cagagagggg acacagcccc 360  
ccctgccctt ggggtcttga gtgtggggg tcagagagag agtgggggac accgccaggc 420  
caggccctga gggcagaggt gatgtctgag tttctgcgtg gccactgtca gtctcctcgc 480  
ctccactcac acag 494

<210> 9  
<211> 865  
<212> DNA  
<213> Homo sapiens

<400> 9  
gtaagggtta cgtgtgatag tctgttccag gatgtgtgtc tctgggatat gaatgtgtct 60  
agaatgcagt cgtgtctgtg atgcgtttct gtgggtggagg tacttccatg atttacacat 120  
ctgtgatatg cgtgtgtggc acgtgtgtgt cgtgggtgat gtatctgttg cgtgcatatt 180  
tgtgtgtgtg tgtgtgtgtg cacgtgtgtg tccatggtgt gtgtgcctgt ggtgtgcatg 240  
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgct gcatgtctgt gatgtgctta 300  
tttgtgtgtg gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat ggggtgtgtg 360  
ggcccccttg ccttactcct tctcctcca ggcattggtc gcaccattgt cctcacgttc 420  
tcgggtgctg gtttggggag ctccacattc agggctcctca ctctagcat ggggtgccct 480  
gtcctgtcac agggctgggc ctgggagact gtaagccagg tttgagagga gaggtaggat 540  
gctgtgtgta ccttccctgga cccctggcac cccagagacc ccagtctggc ctatgccggc 600  
tccatgagat ataggaaggc tgattcaggc ctgcgtcccc gggacacact cctcccagag 660  
cggccggggg ccttggggct cggcaggggt gaaaggggccc ctgggcttgg gttcccaccc 720  
agtgtcatg agcacgctg aggggtaagc cctcaaagtc gtgccaggcc ggggtgcaga 780  
ggtgaagaag tatccctgga gcttcggtct ggggagaggc acatgtggaa acccacaagg 840  
acctctttct ctgacttctt gagct 865

<210> 10  
<211> 3782  
<212> DNA  
<213> Homo sapiens

<400> 10  
tgtgggattg gttttcatgt gtgggatagg tgggatctg tgggattggt ttttatgagt 60  
ggggtaacac agagtccaag gcgagctttc ttctgtagt gggctctgag gtgctccaac 120

agctttattg agggagaccat atcttccttt gaactatggt cgggtttata gtaagtcagg 180  
 ggtgtggagg cctcccctgg gctcccctgt ctgtttcttc cactctgggg tcgtgtgggtg 240  
 cctgctgtgg tgtgtggcgg gtgggcaggg ctccaggcc cctctgtgtt cattggcctg 300  
 gatgtggccc tggctacgct ccgtccttgg aatccccctg cgagtggag gctttctttc 360  
 tttctttttt tctttctttt tttttttttt tgataacaga gtctcgctct tttttgccc 420  
 ggtcggagtg gtttggcgtg atcttggttc actgcaacct gtgcttcctg agttcaagca 480  
 attctcttgc ctacgctccc caagttagctg gaattatagg cgcccaccac catgctgact 540  
 aatttttgta atttttagtag agacgagggt tctccatgtt ggccaggctg gtctcgaact 600  
 cctgacctca ggtgatcctc ccacctcggc tccccaaagt gctgggatga cagggtgtgaa 660  
 ccgcccggcc cggccgagac tcgcttccct cagcttccgt gagatctgca gcgtagactg 720  
 cctgcagcct tgggtgtgac aacctccgtt ttccttctcc aggtctcgct aggggtcttt 780  
 ccatttcatg actctcttca cagaagagtt tcacgtgtgc tgatttccc gctgtttcct 840  
 gcgtaatggt tgtctgctgt ttatcgatgg cctccttcca ttctcttag gctttgttta 900  
 ttgttgtttt tccggctcct tgaaggaaaa gtctcgatta tggatgtttg aactttctct 960  
 tctaaacaag catctgaagt tgcgttttct cctctaaagc agggatcccc agggccctgg 1020  
 ctgtggagtg gcaccggtct ggggctgtgt aggaacccgg cgcacagcgg gaggctagg 1080  
 ggggtgtggg gagccagcgt tcccgcctga gcccgcctcc tctcagatca gcagtggcat 1140  
 gcggtgttca gaggcgacac caccctactg agaactgtgc gtgagagggg tctagattct 1200  
 gtgctcctta tgggaatcta atgctgatg atctgaggtg gaaccgtttg ctcccaaac 1260  
 catccccctc cccactgctg tctgtggaa aaactgtctt ccacgaaacc agtccctgg 1320  
 accacaatgg ttggggaccc tgtgtctaaag acctgttcca gcagcctctc gtcagtgttg 1380  
 atattattgg tttctgtgt tgagtcacaga ataattacgg atttctgtga tgctttccgc 1440  
 cgacctcaga cccatgggct atttggggc gtgttgcctg ctccctgggtt gggaaagggtg 1500  
 caggcccat gtaccttctt gttactgcct tccagggttg ttctcagggt tgaatcgtac 1560  
 tcgatgtggt tttagccac ggcctgcgg ccagctcctg ggggtctggg aacatgtcta 1620  
 agcacagagt caccgtgcgc gtctttgat gcctcacaa gcctcagcct cctgtgtccg 1680  
 tgttagtgtg tgtcacgtgc ctgctcacat cctgtcttgg ggacgcagg gcttagcagg 1740  
 tcccgtagta aatgacaagc gtcttgggg agtctgcaga ataggaggtg ggggtgcgg 1800  
 tctctctccc gcgtcttcag actcttctcc tgctgtgtc gtggctgcac ctgcatcctc 1860  
 gcaatccctc cagcactggg ctggagaggg ccgggagctc gagtgcact tgctgcact 1920  
 gactgtggat ggcagtcggc caggggggtc tgatgtgtg tgactgtgga tggcggttg 1980  
 tcacaggggt ctgatgtgtg gtgactgtg atggcggtcg tgggtctga tgtgtgact 2040  
 gtggatggcg gtcgtgggt ctgatgtgtg gtgactgtg atggcggtcg tgggtctga 2100  
 tgtgtgact gtggtggcg gtgctgggt ctgatgtgtg gactgtggat ggcgtctgtg 2160  
 gggctctgat tgggtgactg ggatggcagt cgtgggtct gatgtgtgt gactgtggat 2220  
 ggcgtctgtg ggtgtctgat tgggtgactgt ggatggcagt cgtgggtct gatgtgtgt 2280  
 gactgtggat ggcgtctgtg ggtctctgat tgtgtgact gtggatggcg gtcgtgggt 2340  
 ctgatgtgtg gtgactgtg atggcggtcg tgggtctga tgtgtgtga ctgtggatgg 2400  
 cgttcgtggg gtctgatgtg gtgactgtg atggcggtcg tgggtctga tgtgtgtga 2460  
 ctgtggatgg tgatcggtca caggggtctg atgtgtgtg actgtggatg gcgtctgtg 2520  
 ggtctgatgt gtggtgactg tggatgtga tcggtcacag ggtctgatg tgtgtgact 2580  
 gtggatggcg gtcgtgggt ctgatgtgtg gtgactgtg atggcggttg gtcgggggg 2640  
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacagg gtctgatgtg tgggtgactgt 2700  
 ggatggcggt cgtgggtctc gatgtgtgtg gactgtggat ggcgtctgtg ggtctgatg 2760  
 tgtgtgact gtggatggcg gtcgtgggt ctgatgtgtg gactgtggat ggcgtctgtg 2820  
 gggctctgat tgggtgactgt ggatggcggt cgtgggtct gatgtgtgt gactgtggat 2880  
 ggcgtctgtt cccgggggtc tgatgtgtg tgactgtgga tggcggtcgt ggggtctgt 2940  
 gtggtgactg tggatggcag tctgtgggtc tgatgtgtg tgactgtgga tggcggtcgt 3000  
 ggggtctgat gtgtgtgtg tgtggatggc ggtcgtggg tctgatgtgt ggtgactgtg 3060  
 gatggcggtc gtgggtctg atgtgtgtg actgtgtgtg gcgtctgtg ggtctgatg 3120  
 ggtgactgtg gatggcggtc gtgggtctg atgtgtgtg actgtgtgtg gtgactgtg 3180  
 acaggggtct gatgtgtgt gactgtggat ggcgtctgtg ggtctgtatg tgtgtgact 3240  
 gtggatggcg gtcgtgggt ctgatgtgt gactgtggat ggcgtctgtg ggtctgtatg 3300  
 tgtgtgact gtggatggcg gtcgtgggt ctgatgtgtg tgactgtgtg atggcagtcg 3360  
 gtcacagggg tctgatgtgt ggtgactgtg gatggcggtc gtgggtctg atgtgtgtg 3420  
 actgtggatg gcgtctgtg ggtctgtatg gtgtgtgtg tggatggcg tctgtgggtc 3480  
 tgatgtgtg tgactgtgga tggcggtcgt ggggtctgat gtgtgtgtg tggatgggtg 3540  
 tcggtcacag ggtctgtatg tgtgtgtgtg gcaggtggag tccaggtgt gtctgtagct 3600  
 actttgcgtc ctcggccccc cggcccccgt tccccaaaca gaagcttccc aggcgtctc 3660  
 tgggtctcat cccgcatcgg ggtctggcg caggtccaca cgtcctgatc ggaagaaaca 3720  
 agtccccagc tctggccggg gcaggccaca tttgtggctc atgcccctc ctctgcgggc 3780  
 ag 3782

12 / 18

<210> 11  
<211> 980  
<212> DNA  
<213> Homo sapiens

5

<400> 11  
gtctgggcac tgccttcgag ggttgggcac ggactcccag cagtgggtcc tccccgggc 60  
aatcactggg ctcacatgacc gacagactgt tggccctggg gggcagtggt ggaatgagc 120  
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggc catgccaggc 180  
10 tgcgacagct gctgcattca ggcacctgct cactgttgac tgcgcgccct ctctccagtt 240  
ccgcagtgcc tttgttcattg atttgctaaa tgtcttctct gccagttttg atcttgaggc 300  
caaaggaaaag gtgtccccct cctttaggag ggcaggccat gtttgagccg tgcctgcgcc 360  
agctggcccc tcagtgtctg gtctgaggcc aaaggaaacg tgtccccctt cttaggagga 420  
cgggcccgtgt ttgagccacg ccccgctgag cgggcctctc agtgcgtggg ctgtccacgt 480  
15 ggccctgttg cccctttgcag atgtgggtctg tccacgtggc cctgtgggtc ttgacagatg 540  
cctgtttagca cttgctcggc tctaggggac agtcgtgtcc accgcatgag gctcagagac 600  
ctctgggcga atttccttgg ctcccagggt ggggggtggg gtggcctggg ctgctgggac 660  
ccagaccctg tgcccggcag ctgggcagca actcctggat cacatatgcc atccgggcca 720  
cgggtgggctg tgtgggtgtg agcccagctg gaccacagg tggccagag gagacgttct 780  
20 ggtcacaca ctctgcctaa gcccattgtg gtctgcagag actcggcccg gccagccac 840  
gatggccctg cattccagcc cagccccgca cttcatcaca aacctgacc ccaaaaggga 900  
cggagggtct tggccacgtg gtctgcctg tctcagcacc caccggctca ctcccatgtg 960  
tctcccgctc gctttcgcag 980

25

<210> 12  
<211> 2485  
<212> DNA  
<213> Homo sapiens

30

<400> 12  
gtgagtcagg tggccagggt ccattgccct gcgggtgggt gggcgggctg gcagggcttc 60  
tgctcacctc tctcctgccc ctccccact gnccttctgc ccggggccac cagagctccc 120  
ttttctggcc cccgccccct ccggctcctg ggctgcaggc tcccagggcc ccggaaacat 180  
ggctcgggct gcggcagccg gagcgagca ggtgccacac gaggcctgga aatggcaagc 240  
35 ggggtgtgga gttgctcctg cgtggaggac gaggggcggg ggggtgtgtc gggtcagggt 300  
tgcgccgagc gtttgagcct gcagcttgtc agctccaaat tactactgac gctggacacc 360  
cggctctcac acgcttgtat ctctctctcc cgatacaaaa ggattttatc cgattcccat 420  
tctgtccct gtctgtgtac ccccgcgagg gcgcgggctc ttctctctgt gactagattt 480  
cccatctgga aagtgcgggg ttgaccgtgt agtttgcctc tctcgggggg cctgtgtgtg 540  
40 ccatggggca ggcggcctgg gagagctgcc gtcacacagc cactgggtga gccacactca 600  
cgggtgtaga gccacagtgc ctgggtccac atcacgtcct ctggatttta agtaaaacca 660  
cacacctccc ggcaggcacc tgcttgcgac cctgtgtgtg cctggggaga gtggtagcac 720  
ggaggaatt cgtgcacact caaggtcatc agcaaggcca tccgcagcca ggtggaacgt 780  
ggaggcctct ctctgggagc gtctccagcg gataaaggac tgtgcacagc ttcggaaagct 840  
45 tttatttaaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900  
attataatat ttattaaagt ataattagaa atattaagta gtacacacgt tctggaaaaa 960  
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020  
taagcggccc ccaggccccc agaattcgct gacaaagtca cctccccaga gaagccacca 1080  
cgggctcctc tcgtggtcgt gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140  
50 gtggcagggg tttggggaat gtgaggtgat gactgcgtcc tcatgccctg acagacagga 1200  
gggtgactgt tctgtcctgt ccctaggaca cggacaggcc cgaagctcta gtccccatcg 1260  
tggtccagtt tggcctctga ataaaaacgt cttcaaaacc tgttgcccc aaaaactaaga 1320  
acagagagag tttcccatcc catgtgctca caggggcgta tctgcttgcg ttgactcgct 1380  
gggctggcgg gactcctaga gttggtgcgt gtgcttctgt gcaaaaagtg cagtcctctt 1440  
55 gcccacact gtgatatctg caccagcaag gaaagcctct tttcttttct tctttttttt 1500  
ttttttgaga cggaaagcga ctgttgcctg cctgggcttg agtgacgtgg cgcgactctca 1560  
actcactgca acctccgctt cccgggttcc agcattttct ctgcctcagc ctcccagca 1620  
gctgagatta caggcaccca cccctgcgc ctggctaatt tttgtatttt tagtagagag 1680  
gggtttttgc catgttggcc aggtgtgtct cgaactcctg acctcaggtg atccaccac 1740  
60 ctccggcctcc caaagtgtg gattacagg tgtgagccat caccggcagc cggaaagcct 1800  
ctttttaagg tgaccacctg tagcgcttcc cgaataaac aggtcttctt tttgcagtag 1860  
gctgcaagcg tctcttagca acaggagtgg cgtcctgtgg gctcgggga tggctgaggg 1920  
tcgctgggca gccatgcctt ctgtgtgcac ctttaggttc caccgggcta tctgtctctc 1980  
actgtttgtc tgaacacgca ccttggcat cctgttttgg agagtctctg cttctcgttg 2040  
65 gtcagctga aactaggggc aaggtgtgat ccgttggcgc gcagcggcta catgtagggt 2100

13 / 18

catgagtcctt tcaccgtgga caaatccctt gaaaaaaaaa aaaggagtc cc ggtaagcat 2160  
 tcattccggg tcaagtgtct ggttctgtga ataaactcta agatttaaga aaccttaatg 2220  
 aaagaaaacc ttgatgattc agagcaagga tgtggtcaca cctgtggctg gatctgttcc 2280  
 agccgccccca gtgcatgggt agagtgggga gcaggagattg tttgttcaga ggtctcatct 2340  
 ggatgttttc tgagggtgtt gccggctgaa tggtagacgt gtcgtttgtg tgtatgaggt 2400  
 tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgccccgc 2460  
 tctcacctgt gtcttccgc cccag 2485

<210> 13  
 <211> 1984  
 <212> DNA  
 <213> Homo sapiens

<400> 13  
 gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tcatgcatcc 60  
 agtggttaata ttcttggtgc tctggagacc atgactgctc tgtcttgagg aaccagacaa 120  
 ggttgacgcc ccttcttggg atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180  
 gggctccacg caggctctgt ccagcggcca tgtccagagg cctcagggct cagcaggcgg 240  
 gaggcccgct gccctgcatg atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300  
 ctgtcacgtc acccaggctc cgttagggtc cttggggaga tggggctggg gcagctgag 360  
 gccccacatc tcccagcagg ccttcgacag gtggcctgga ctgggcgcct cttcagccca 420  
 ttgccatcc cacttgcatg ggttctacac ccaaggacgc acacacctaa atatcggtcc 480  
 aacctaatgt ggttcaactc agctggcttt tattgacagc agttactttt tttttttaa 540  
 tactttaagt tctaggttac atgtgcacga cgtgcagggt agttacatat gtatacatgt 600  
 gccatgttgg tgtgctgcac ccattaactc atcatttaca ttaggtatat ctccaatgac 660  
 tateccctcc cactcccccc atcccatgac aggccttggg tgtgtatgtt ccccacctg 720  
 tgtccaagtg ttctcattgt tcagttccca cctgtgagtg agaacatgtg gtgtttgggt 780  
 ttctttcctt gcaatagttt gctcagagtg atggtttcca gcttcgtcca tgtccctaca 840  
 aaggacatga actcatcctt ttttatgact gcatagtatt ccgtgggtga tatgtgccac 900  
 attttcttaa tccagtctat catcgatgga catttggtt ggttgcaagt ctttgctact 960  
 gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020  
 tcctttgggt atatacccg taatgggatg gctgggtcaa atggtaattc tagttctaga 1080  
 tccttgagga atcaccacac tgtcttccac aatggttgaa ctagtttaca ctcccaccaa 1140  
 cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt ttatgaaaa 1200  
 tagtatcact gaacaagcag acagttagtg aaggatgcgt caggaagcct gcaggccaca 1260  
 cagccatttc tctcgaagac tccgggtttt tctgtgcat cttttgaaac tctagctcca 1320  
 attatagcat gtacagtgga tcaaggctct tctcattaa ggttcaagt ctagattgaa 1380  
 ataagttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttga 1440  
 ggaaagtgtc ctgcagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500  
 ccattggtcat ggggcgctgg gcttgggcct gaggttcaca cagtgcacca tgcccagctt 1560  
 cctgtggata ggatctgggt ctcgatcat cctggagacc acagctgcca tgcgtgtaaa 1620  
 gggcaccacg tggctcagag gggcgaggt tcccagcccc agctttctta ccgtcttcag 1680  
 ttatttttcc ctaagagtct gagaagtggg gccgcgctg atggccttcg ttcgtcttca 1740  
 gctggcacag aattgcacaa gctgatggta aacactgagt acttataatg aatgaggaat 1800  
 tgcgttagca gtttaactgta gagagctcgt ctgttggaag gaaatttaag tttttctatt 1860  
 aaccgctttg gagaatgtta ctttatttat ggctgtgtaa attgtttgac attcagctccc 1920  
 tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcccttatt 1980  
 ttag 1984

<210> 14  
 <211> 1871  
 <212> DNA  
 <213> Homo sapiens

<400> 14  
 gtgaggcccg tgccgtgtgt ctgtggggac ctccacagcc tgtgggcttt gcagttgagc 60  
 ccccgtgtc ctgccccctg caccgcagcg ttgtctctgc caagtcctct ctctctgccc 120  
 gtgctggatc cgcaagagca gaggcgcttg gccgtgcacc caggcctggg ggcgcagggg 180  
 caccctcggg agggagtggg taccgtgcag gccctggctc tgcaagagacg caccaggtt 240  
 acacacgtgg tgagtgcagg cgggtgacctg gctcctgctg ctctttggaa agtcaagagt 300  
 ggccgctcct ggggccccag tgagaccccc agggagctgtg cacaggccct gcagggccga 360  
 ggcgcgagcc tcttccccag ggtgcacctg agcctgcgga gaggcaggag tgcgtagtga 420  
 gctggcccac agcgttcgct gcggtcacgt tccctgcgtg ggttgtttgg gatcggtggg 480  
 agaatttggg ttgtctgagt gctgctgtct tgaaccacgg agatggctag gagggggttt 540  
 cagagttgat ttttgtgaat caaactaaaa tcaggcacag gggacctggc ctacgcacag 600

5 gggattgtcc aatgtgtgcc ccctcaaggc cgccccacag agccggtggg cttgttttaa 660  
 agtgcgattt gacgaggagc gagaaacctt gaaagctgta aagggaaacc tcagaaaaatg 720  
 tggccgcccag ggggtgtttc aggtgctttg ctgggctgtg tttgtgaaaa ccattttgga 780  
 cccgccctcc aagtcacccc tccaggtcca cctccaggc cgcctctggg ctgggggtat 840  
 gcctggcggt ccttgtgccg cagcccgagc cacagcaggc tgtgcacatt taaatccact 900  
 aagattcact cggggggagc ccaggtccca agcaactgag ggctcaggag tctcagggtc 960  
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagtctc tgagggtgct 1020  
 ggccaggagg gtggctcaga gtgtatgttg ggtcccacc gggggcagaa ctctgtctct 1080  
 gatgagtcgg cagccatgta acaggaaggc gtggccacag ggagctggga atgcaccagg 1140  
 10 ggagctgcgc agctggccga ggtcccaggc ccaggccaca gaaggggcag ggggacgccc 1200  
 gggggccacag cagaggccgc aggaagggaa ggggatgccc aggccagagc agaggctacc 1260  
 gggcacaggg gggctccctg agctgggtga gcgaggctca tgactcggcg agggaaacct 1320  
 cttgacgtga agctgacgac tgggtgtgcc cagctcacag ccagccaggc tcccgcgcct 1380  
 gagcaggaaac tcagaacctt cccctttgtc taaagcacag cagatgcctt cagggcatct 1440  
 15 aggagaaaaac aggcagaatc gttgagaaac gctttaaag aaggtgggat ggtggcaatt 1500  
 tctgtccag attttagtct gcccggacc acagatgagt ctataacggg attgtggtgt 1560  
 tgccatgggg acacatgaga tggaccatca cagaggccac tggggctgca cctcccatct 1620  
 gagtccctggc tgtcccggtt ccaggccagg tcttctgcat ctcacctacc tgcctgccc 1680  
 gggagacagg gaaagcaccc cgaagtctgg agcagggtc ggccaggct cctcagagct 1740  
 20 cctgccaggc ccagcacctt gctccaaac accactctc tggggttttc caaagcattt 1800  
 aacaagggtg tcagggtacc tctgggtga cggccccgca tctggggct gacatgccc 1860  
 ctctgcctta g 1871

<210> 15  
 <211> 3801  
 <212> DNA  
 <213> Homo sapiens

30 <400> 15  
 gtgagcgac ctggccggaa gtggagcctg tggccggctg gggcagggtg tgctgcaggg 60  
 ccgttgctgc caccctctgt tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120  
 ggccacaggg tggccctcgt cccatctggg gctgagcaga aatgcattct tctgtgggag 180  
 tgagggtgct cacaacggga gcagttttct gtgctatttt ggtaaaagga aatggtgcac 240  
 35 cagacctggg tgactcaggc tgtcttcaga aagcagctct gatccgaacc caagacgccc 300  
 gggccctgct gggcgtgagt ctctcaaac cgaacacagg ggcctctgtg ggcattgagc 360  
 cctctgaacc cgagaccctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420  
 caggggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaaag gctcatccac 480  
 agtctacagg atgccatgag ttcattgatc cgtgtgaccc atcaggggac agggccatgg 540  
 tgtggggggg gtctctacaa aattctgggg tcttgtttcc ccagagcccc agagctcaag 600  
 40 gcccgcctc aggcctcagc acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660  
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720  
 ataatcccag cactttggga ggccgagggt ggtggatcac ttgaggccag gagtttgagg 780  
 ccaacctaac caacatagtg aaattccatt tctacttaaa aaatacaaaa attagccttg 840  
 cctggtggca cagcctgta gtccccgcta tggcggaggc tgaggcagga gaatcatttg 900  
 45 aaccaggag gcagagggtg cagtgcagcc agatcacacc actgcactcc agcctgggca 960  
 acagagttag acttcatctt aaaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020  
 ggacagggtg ttttttatcc tgtccttcga taatatttac tgggtgctgtg ctaggaggct 1080  
 gaactggggg tgccttcctc tgaaaggcac accttcattg gaagagaaat aagtgggtga 1140  
 tgggtgttaa accagagggt taaactgggg tcctgtcgtt ctgagttaac agtccagatc 1200  
 50 tggactttgc ctctttccag aatgctccct ggggtttgct tcatggggga gcagcagggt 1260  
 tggacacctt cgtgatgggg gagcagcagg tgcagacgcc ctcatgatgg gggagtggca 1320  
 ggtgcagaca cccttgtgca tgggtcccag catgtccctg ttgcagctcc cccccacaa 1380  
 ggatgcgggt ctctgtgct ccccacagtc cctgcttccc tctcacagcc ttacctggct 1440  
 ctggcctcca ctggcttgt ctgcattgct tccacatttc ctgggctccc agcacctctt 1500  
 55 cgctctccc aggcacctct gcagtgctgg ccataccagt cagctgtgaa ctgtccactg 1560  
 cttattttgc tccccatgaa atgtattttt taggacaggc acccctgggt ccagcctctg 1620  
 gcacagcacc agtgaatgtt attgaaggac aaaggacaga caaacaatc agggaaatgg 1680  
 gttctctcta aacacattgc aaagccacag aggcctagtgc aggatgggtg ggcatcagg 1740  
 catcagatgt ggggtccaat ccagaatatt ctgtgctccc aaaggccact tggtcagagt 1800  
 60 gtgtgcttgc agagggtgct ctaaaagctc agcagtggag gcagtgttcc gccatactca 1860  
 ggggtgaactc acatcctctg tgtctgaagt atacagcaga ggcctgaagg gcatctggga 1920  
 gaagaaaaca ggcaaaatga ttaagaaaag tgaaaaagga aaagtggtaa gatgggaatt 1980  
 ttcttgtcca gattttagtc tcccaaacca cagctcagat ggtagaatgt ggtcagaact 2040  
 gatggacaga acaatagaac aaaacggaag ccctatctct cagaaacgtg tgttaattgt 2100  
 65 gtatgtggca cagctgatgg aaaaagagat gtgtgtgtaa ttttttttcc tgagaaaact 2160

15 / 18

5 gactggaagc aaataagttg tgtctttaca gcatatacca gagcagattc taggtagaag 2220  
 aggagacaca tgcaacaac accagcaaca gaaataaac aaaagactca aaggggaagg 2280  
 aggtgaacgt tccctgggtt gggtgtgggg aaggacacac agggaggcgg atgaaaccag 2340  
 tgaggcaacg ggcattgctt tcaactgcaga gaaactcagc ttgcctgagc cacagtgaag 2400  
 atggccattc cctggagcgt ttgtgcacgt gatttattta aggcgccctg tgaggctcctg 2460  
 cacattcatc ctctcacttt gttctcctaa ccacctgaga ggtagaggag gaaaggctcc 2520  
 agggggagcag ccgcctcttg tcacccagct ggcaaaaggg atgcatgatt gcagcctggc 2580  
 ctcttgctcc ggggcccttg ctctgccga ggacccaca caagtcagac ccataggctc 2640  
 aggggtgagcc ggagcccaag gtctgtgttg ggatggctgt gaaagaagaa atggagctct 2700  
 gatgcacact tgggaaggtc ctaccagcag cgtcaagaa atgcatgtga aactgacagc 2760  
 gagaccatc cctcaaagaa acgcacgtga aactgatggc gagacctgtc cccatccctc 2820  
 atgctggctc cttttctggg cttgccaaaga gccagcatca ggttgaggca agctggaaag 2880  
 acttttctgg aaagcagctt gtttgcattg aagtcctcac aatgtcctgt gtctcccg 2940  
 taattccact tctgaagtga ccagacatta tcacgggtct tatttaccat tccagtggt 3000  
 15 ccaggcaggg ggacttgcca cagcaagtca cgaacctgcc caaatacagg gctaaggaga 3060  
 tattatgcat cacaaaactt gctctgccat taaacatttt tcaaagaatt ttgagaat 3120  
 gtttaatggc acaaaactt tattcaatg tagcagtgtt caaagctgga tgaagaa 3180  
 cacacccag gagcctgccc tgaatgtcat gtgtgttcat ctttgacat ggacatacat 3240  
 20 gggcagtgag tgggtggtag gccctggagg acatcggtgg gatgcctcca tcctgcccct 3300  
 ctggagacac catgtgtgcc acgtgcactc actggagccc tgtttagctg gtgccacctg 3360  
 gctcttccat ccctgagatt caaacacagt gagattcccc acgcccaact cagtgtcttc 3420  
 ccacaaaaaa cctgagtcac acctgtgttc actcgaggga cgcccgggag ccagggtctc 3480  
 acagtttatt atgtgttttt ggctgagtta tgtgcagatc tcatcagggc agatgatgag 3540  
 25 tgcacaaaca cggcgtgctg aggtttggat acactcaaca tcaactagcca ggtcctggtg 3600  
 gagtttggtc atgcagagtc tggatggcat gtagcatttg ggtccattgg agtgagcacc 3660  
 cagcccccct gggctgcagc gcatgcccga ggcaggacaa ggaagcggga ggaaggcagg 3720  
 aggtctcttg gagcaagctt tgcaggaggg ggctgggtgt ggggcaggga cctgtgtctg 3780  
 acattcccc ctgtgtctca g 3801

30 <210> 16  
 <211> 880  
 <212> DNA  
 <213> Homo sapiens

35 <400> 16  
 gtgagcaggc tgatggtcag cacagagttc agagtccagg aggtgtgtgc gcaagtatgt 60  
 gtgtgtgtgt gtgcgcgctg gcctgcaagg ctgatggtga ctggctgcac gtaagagtgc 120  
 acatgtacgc atatacacgt gagcacatac atgtgtgcat gtgtgtacat gaaggcatgg 180  
 40 cagtgtgtgc acaggtgtgc aagggcacaa gtgtgtgcat atgcgaatgc acacctgaca 240  
 tgcattgtgt ttctgtgcaca gtctgtgtgg cattcacgtg aggtgcatgc gtgtgggtgt 300  
 gcagtgtgag tagcatgtgt gcacataaca tgtattgagg ggtcctcgtg ttcaccccgc 360  
 taggtcctca gcaccagtgc cactccttac aggatgagac ggggtccagc gccttggtgg 420  
 gctgaggctc tgaagctgca gccctgaggg cattgtccca tctgggcatc cgcgtccact 480  
 45 cctctcctc tgggtctctg tgtccactcc cctctcctg tgggcattta catccactcc 540  
 actcctctc tcctgtgggc atccgcgtcc actccccctc tctgtgggca tctgcgtcca 600  
 cctccccct ctgtgggcat ttgcgtccac tccctctcct ggttctctcc tgtcttggtc 660  
 gagcctcggg ggcaggcaga tgacacagag tcttgactcg ccagggtgg ttcgcagctg 720  
 cgggtgagg gccaggccgg atttcactgg gaaggaggat agtttcttgt caaatgttc 780  
 50 ctctttcttg ttccatctga atggatgata aagcaaaaag taaaaactta aaatcccaga 840  
 gaggtttcta ccgtttctca ctctttcttg gcgactctag 880

55 <210> 17  
 <211> 3186  
 <212> DNA  
 <213> Homo sapiens

60 <400> 17  
 gtgagccgcc accaaggggt gcaggcccag cctccaggga cctccgcgc tctgtctacc 60  
 tctgaccggg ggcttcacct tggaaactcct ggggttttagg ggcaaggaaat gtcttacgtt 120  
 ttcagtgggt ctgctgcctg tgcacagttc tgttcgcgtg gctctgtgca aagcacctgt 180  
 tctccatctc tgggtagtgg taggagccgg tgtggcccca ggtgtcccca ctgtgcctgt 240  
 gcactggcgg tgggacgtca tggaggccat cccaggggcag caggggcatg gggtaagag 300  
 atgtttatgg ggagctctag cagaggaggc tgggaagggt tctgaacagt agatgggaga 360  
 65 tcagatgccc ggaggatttg gggctctcagc aaagagggcc gaggtgggtg cagggtgagg 420  
 tcgctggccc caccggggg aaggtgcagc agactgtgtg ctccccacac agcccgcca 480

16 / 18

5 gcacctgtgc tctgggcatg gctgtgctcc tggaaacgttc cctgtcctgg ctggtcaggg 540  
 ggtgccccctg ccaagaatcg acaactttat cacagaggga agggccaatc tgtggaggcc 600  
 acaggggccag ctctctgccc gactcagggc aggtgggtggc acaagcctcg gggctgtacc 660  
 aaagggcagt cgggcaccac agggccgggc ctccacctca acaggcctcc cgagccactg 720  
 ggagctgaat gccaggaggc cgaagccctc gccccatgag ggctgagaag gactgtgagc 780  
 atttgtgtta cccaggggcg aggtgcgcgc aattaccgtg cacacttgat gtgaaatgag 840  
 gtcgtcgtct atcgtggaat cccagcaagg gctcacggga gacttttcca ttacaaggtc 900  
 gtacccatgaa aatgggtttt aacccgagtg cttgcgcctt catgctctgg caggggaggc 960  
 10 agagccacag ctgcatgtta ccgcctttgc accagctcca gaggcttggg accaggctgt 1020  
 ctgagttcca ggggtgctcc ggctcagacc gccctcctct ctgccttctc tctctgcctc 1080  
 aaatcttccc tctgttgcac ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140  
 ctcccttccg gaaacccttg ggtgtgctg gatacagggt cactgagga ctggagggtg 1200  
 ctgacactgt ggttgacccc aggggtccagc tggcgtgctt ggggcctcct tggggccatga 1260  
 tgaggtcaga ggaagttttc caggtgaaaa ctccctggga actcccaggg ccagtgtacc 1320  
 15 tgccacctgc tcccccata ttcagctcag tcttgcctc atttcccac cagggtctct 1380  
 agctccgagg agctcccgta gagggcctgg gctcaggga gggcggctga gtttcccac 1440  
 ccagtgggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatcgga 1500  
 tggggccacg gccgtttcca aacacagagt caggcacgtg gaaggccacg gaatccccct 1560  
 ccttcgaggc aggagtgagg gaacggagag ctggggcccg atttcacgc agccaggctg 1620  
 20 cagtgggga ggtgtggtg gtccacgtgg cgtggggggc ggggtctgat tcaaatccgc 1680  
 tggggctcgg ccttccctgg ccgtgctggc cgcgcctcca cacgggcttg ggggtgacgc 1740  
 cccgacctct agcagggtggc tatttctccc ttggaagag agccctcac ccagtctagg 1800  
 tgtttccctc ctgggtcagg agcgtggccg tgtggcaacc cggggacctt aggtctattt 1860  
 atttgtttaa aaacattctg gccctggctt ccgttgttgc taaatgggga aaagacatcc 1920  
 25 cactcagca gacttactga gaggctgaaa ccggggtgct ggcttgactg gtgtgatctc 1980  
 aggtcattcc agaagtggct caggaaagtc gtgagaccag gtacatgggg ggctcaggca 2040  
 gtgggtgaga tgaggtagac ggggggctca ggcagtgggt gaggccagggt acatgggggg 2100  
 ctcaggcact ggttgagatg aggtacacgg ggggctcagg cagagggtca gaccaggtag 2160  
 acggggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcatggtag 2220  
 30 ccaggctctg gcacacctgc cccaaagtcc caggaaagct agaggccaaa gatggaggct 2280  
 gacagggtcg gcgcggtggc tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340  
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400  
 tatgaaaaat aaaaacaaaa attagtgaa catggtgggt tgccgctgta gttccaatca 2460  
 35 ttgggaggct gaagtgggag gatcacttga gccaggagg tggaaagctg agtgagctga 2520  
 gattgcacca ctgtactgca gcctgggtga cagagtga gcccattctc acaacaacaa 2580  
 agaagactga caaatgcagt ttcttggaaa gaaacattta gtaggaaactt aacctacaca 2640  
 cagaagccaa gtcggtgtct cgggtgctag gagatgagat gatgggtcct cacaccatca 2700  
 cccagaccc aggggtttatg caccacaggg gcgggtgggt cagaagggat gcgcaggacg 2760  
 ttgatatacg atgacatcaa ggttgtctga cgaagggcag gattcatgat aagtacctgc 2820  
 40 ttgtacacaa ggaacaatgg ataaactgga aaccttagag gccttcccgg aacaggggct 2880  
 aatcagaagc cagcatgggg ggctggcatc caggatggag ctgcttcagc ctccacatgc 2940  
 gtgttcacac agatgggtgca cagaacgcga gtgtacctgt gcacacacag acacgcagct 3000  
 actcgacac acaagcacac acacagacat gcatgcatgc atccgtgtgt gtgcacctgt 3060  
 gcccattgag aaacccatgc atgtgcatc atgcacgcac acaggcaccg gtggggccat 3120  
 45 gcccacaccc acgagcaccg tctgattagg aggcctttcc tctgacgctg tccgccatcc 3180  
 tctcag 3186

<210> 18  
 <211> 781  
 50 <212> DNA  
 <213> Homo sapiens

<400> 18  
 55 gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60  
 ggagactgag tgaatctggg ctttaggaagt tcttaccctt ttctcgatca ggaagtgggt 120  
 taacccaacc actgtcaggc tctgtctgcc gccctctcgt ggggtgagca gagcacctga 180  
 tggaaaggac aggagctgtc tgggagctgc catcttccc accttgcctt gcctggggaa 240  
 gcgctggggg gcctgggttc tctgttttgc cccatgggtg gatttggggg gcctggcctc 300  
 60 tctgttttgc cctgtgtgtg gattgggctg tctccctgct atggcacctta gggcccttgt 360  
 gcaaacccag gccaaaggct taggaggagg ccaggccacg gctacccccc cctctcagg 420  
 agcagaggcc gcgtatcacc acgacagagc cccgcggcgt cctctgcttc ccagtcaccg 480  
 tctctgtccc ctggacactt tgtccagcat cagggagggt tctgatccgt ctgaaattca 540  
 agccatgtcg aacctgcggc cctgagctta acagcttcta ctttctgttc tttctgtgtt 600

17 / 18

gtggaaatctt cacctggaga agccgaagaa aacatttctg tctgtactcc tgcggtgctt 660  
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccg 720  
gtgtctcctg ggaggggagc tgggctgggc ctgtgactcc tcagcctctg ttttccccca 780  
g 781

5  
<210> 19  
<211> 536  
<212> DNA  
<213> Homo sapiens

10  
<400> 19  
gcaagtgtgg gtggaggcca gtgcggggccc cacctgcccc ggggtcatcc ttgaacgccc 60  
tgtgtggggc gagcagcctc agatgctgct gaagtgcaga cgcccccggt cctgaccctg 120  
ggggccttga gccacgctgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180  
15 agcctggcag ggtccccaac ttcttgaacc cctgttcccc atctcagggg cgatggctcc 240  
ccacgcttgg gagccttctg acccctgacc tgtgtcctct cacagcctct tccctggctg 300  
ctgccctgag ctcttggggt cctgagcaag ttctctcccc gcccgccgc tccagcgcca 360  
ctgggctgct tgtctgctcg ccccggtgga ggggtgtctg tcccttcaat gaggttcccc 420  
ccagccaggg ccacgaggtg caggccctgc ctgccggcc acccacacgt cctaggaggg 480  
20 ttggaggatg ccacctctgg cctcttctgg aacggagtct gattttggcc ccgcag 536

<210> 20  
<211> 3179  
<212> DNA  
<213> Homo sapiens

25  
<400> 20  
atctcatgtt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60  
ctgtgagtga acggggtggg ggtcagtgct ggcccatggc ctggctgtgc atttacggaa 120  
30 gtctatgagt gaatggggtt gtggtcagtg cgggcccatt gcctggctgg gcctgggagg 180  
tttctgatgc tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagccccc 240  
ccctggaaga cataacagta agtccaggcc cgaaggggcag cagggatgct gggggcccg 300  
cttgggcccgc ggggatgatg gagggcctgg ccagggtggc agggatgatg ggggccccag 360  
ctgggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420  
35 ggctggggccc cctcctcccc tgcctccac ctgcagccgt ggatccggat gtgcttccct 480  
ggtgcacatc ctctgggcca tcagctttca tggagggtgg gggcaggggc atgacacat 540  
cctgtataaa atccaggatt cctcctcctg aacgccccaa ctcagggtga aagtcacatt 600  
ccgctcttgg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660  
gtggggcagt ggagggtgtg gacacaggag gcttcagggt ggggctgtgt atgctctctc 720  
40 atcctcttat catctccag tctcatctct catctctta tcatctccca gtctcatctg 780  
tcttctctt atctccagct ctcatctgtc atcctcttac catctccag tctcatctct 840  
tactctctta tctcctagtc tcatccagac ttactctcca gggcgggtgc caggctcgca 900  
gtggagctgg acatacgtcc ttctcaggc agaaggaaat ggaaggattg cagagaacag 960  
gagggggcgc tcagagggac gcagctctgg ggtgaagaaa cagccctccc tcagaagtgt 1020  
45 gcttgggcca cagaaaaccg agggccctgc gtgagtggtc ccagagcctt ccagcaggtc 1080  
cctggtgggg ccttatggta tggccgggtc ctactgagtg cactctggac agggcttctg 1140  
gtttgagtgc agcccgagc tgcctgggtg cggggtgggg gcttatggcc actggatatg 1200  
gcgtcattta ttgctgctgc ttcagagaat gtctgagtga ccgagcctaa tgtgtatggt 1260  
gggcccaggt ccacagactg tgtcgtaaat gcactctggg gcctggagcc cccgtatagg 1320  
50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaat 1380  
ggaaggggagc ggccccgggc gccgtgggcg gacgacctca agtgagaggt tggacagaa 1440  
agggcgggga cttcccagga gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500  
gaccaacagg tcaggccatt gttcagctat ccattctcta caaagctcca gattcctgtt 1560  
tctccgggtg ttttttgggt aaattttact caggattact tatatttttt gctaaagtat 1620  
55 tagaccctta aaaaagggtat ttgctttgat atggcttaac tactaagca cctactttat 1680  
ttgtctgttt ttatttatta ttattattat tattagagat ggtgtctact ctgtcaccca 1740  
ggtgtgtagt gcagtggcac agtcatggct cgtgtgagcc gcaaaacccc aggtccaagt 1800  
gatcctccgg cctcagcttc ccagagtgtc gggattacag gtgtgagcca ctgcccttgc 1860  
ctggcacttt taaaaaccac tatgtaaggt cagggtccagt ggcttccaca cctgtcatcc 1920  
60 cagtagtttg ggaagccgag gcagaaggat tgtctgaggg caggagtttg agaccagcat 1980  
gggtaacata gggagacccc atctctacaa aaaatgcaaa aagttatccg ggcgtggggt 2040  
ccagcatctg tagtcccagc tgctcgggag gctgagtggt aggatcgctt gagcccgga 2100  
ggtcatggct gcagtggact gtgattgtac catcgactc cagcctgggc aacagagtga 2160  
gacctgtct caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaggaa 2220  
65 gaaggaaaga gaagaagaag gaagaaggaa gaaagaagga gaaggaggcc tgctaggtgc 2280



taggtagact gtcaaatctc agagcaaat gaaaataaca aagttttaaa gggaaagaaa 2340  
aaccacagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttccaca 2400  
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460  
ggaggctgtg ggtgacacca gccaggaccc ctgaaagga gtggttggtt tcctgcctca 2520  
5 gccccacgct cctgccggtc ctgcacctgc tgaaccgtc gatgttggtg ccagggtgcc 2580  
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagcccag 2640  
gcacttgtgg caggcacaat tacagccctt ccccaaagat gcccacgtcc ttctcctgga 2700  
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcagggtg gaatcacggc 2760  
tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820  
10 ggggtccctag aagtgcagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880  
cactggccac tgctggcttt gagatggagg agggggctcc cagccaagga atgggggag 2940  
ccgtccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgcc 3000  
acgcctcgat ttcaggccag tgggacctgt ttcagctttc cgccctccag agctgtaaga 3060  
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120  
15 cagtacaggg aaatgaatac agggacagtt ctcagagtga ctctcagccc acccctggg 3179